

凸分析及应用捷径

〔美〕Boris S. Mordukhovich, Nguyen Mau Nam　著
赵亚莉　王炳武　译

科 学 出 版 社

现代数学译丛 27

凸分析及应用捷径

〔美〕Boris S. Mordukhovich, Nguyen Mau Nam 著

赵亚莉 王炳武 译

科学出版社

北京

图字：01-2015-5618

内 容 简 介

凸最优化在数学、应用科学和实际应用的许多领域中的影响日益增长. 现在许多大学正讲授它，而且被不同领域的研究人员应用. 由于凸分析是凸最优化的数学基础，深入的凸分析知识可帮助学生和研究人员更有效地利用其中的工具. 本书的主要目的是提供一个容易进入到凸分析及其在最优化中应用的最基础部分. 变分分析的现代技术被用来阐明和简化凸分析中的一些基本证明，并且在有限维空间中建立凸函数和凸集的广义微分理论. 我们还给出凸分析在选址问题以及许多令人感兴趣的几何问题，如 Fermat-Torricelli 问题、Heron 问题、Sylvester 问题及其推广中的新应用. 当然，我们不期望触及凸分析的每个方面，但是对这个学科的初级教程来说本书包含足够的素材. 它也可作为凸最优化及应用课程的补充阅读材料.

本书可作为高年级本科生及研究生凸分析及其应用课程的教科书. 也可供相关专业科研人员参考.

图书在版编目(CIP)数据

凸分析及应用捷径/(美)莫尔杜霍维奇(Mordukhovich, B.S.)等著；赵亚莉，王炳武译. —北京：科学出版社，2015.9
(现代数学译丛)
书名原文：An Easy Path to Convex Analysis and Applicantions
ISBN 978-7-03-045654-0

Ⅰ. ①凸… Ⅱ. ①莫… ②赵… ③王… Ⅲ. ①凸分析 Ⅳ. ①O174.13

中国版本图书馆 CIP 数据核字(2015) 第 216527 号

责任编辑：李 欣 / 责任校对：钟 洋
责任印制：张 伟 / 封面设计：耕者设计

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

北京教图印刷有限公司 印刷
科学出版社发行 各地新华书店经销
*
2015 年 9 月第 一 版 开本：720×1000 B5
2015 年 9 月第一次印刷 印张：12 1/4
字数：242 000

定价：78.00 元

(如有印装质量问题，我社负责调换)

原著第一作者以此书纪念他的父亲 Sholim Mordukhovich (1924-1993)：一位善良的人，一位勇士.

原著第二作者以此书纪念他的父母.

作 者 简 介

Boris S. Mordukhovich (莫尔杜霍维奇) 是州立韦恩大学的数学“大学精英教授”(Distinguished University Professor), 也是法赫德国王石油矿产大学数学与统计学的讲席教授. 他已经发表了 350 多篇文章并出版数部专著. 他最著名的成就是强有力广义微分结构的引入和发展及其在变分分析、最优化、均衡、控制、经济、工程以及其他领域的许多类问题中的应用. Mordukhovich 是 SIAM 院士, AMS 院士和许多国际奖以及荣誉的获得者, 包括世界上六所大学授予的荣誉博士学位. 他被 ISI(Institute of Scientific Information) 评为数学方面高引用学者. 他的研究连续受到美国国家自然科学基金的资助. 地址：美国密歇根州底特律市州立韦恩大学数学系. 邮编：48202. 电子邮件：boris@math.wayne.edu. 网页: math.wayne.edu/~boris.

Nguyen Mau Nam(阮茂南) 于 1998 年在越南顺化大学获得理学学士学位, 于 2007 年在州立韦恩大学获得理学博士学位. 他目前是州立波特兰大学的数学助理教授. 他已有大约 30 篇关于凸分析和变分分析、最优化及其应用的论文被高水平的数学杂志发表或接收. 他的研究得到西蒙斯基金会合作项目资助. 地址：美国俄勒冈州波特兰市州立波特兰大学马席赫数学与统计系. 邮编：97201. 电子邮件：mau.nam.nguyen@pdx.edu. 网页: web.pdx.edu/~mnn3.

关键词: 仿射集, Carathéodory 定理, 凸函数, 凸集, 方向导数, 距离函数, Fenchel 共轭, Fermat-Torricelli 问题, 广义微分, Helly 定理, 极小时间函数, Nash 均衡, 法锥, Radon 定理, 最优值函数, 最优化, 最小封闭圆问题, 集值映射, 次微分, 次梯度, 次梯度算法, 支撑函数, Weiszfeld 算法.

译　者　序

本书是关于凸分析的一本新著, 原著作者都是变分分析领域的著名学者, 其中 Boris S. Mordukhovich 是现代变分分析的创始人之一. 本书涵盖了有限维空间上凸分析的一些基本结果, 并涉及了凸分析的一些最新应用; 现代变分分析的思想方法融入了整个理论的阐释与处理, 因此本书不仅是凸分析的入门教材, 也为读者进一步进入变分分析领域打下了基础.

本书译者曾经参与翻译 Mordukhovich 教授的专著《变分分析与广义微分》上、下卷, 分别于 2011 年与 2014 年由科学出版社出版. 在 Mordukhovich 教授和王炳武教授的支持下, 赵亚莉完成译本的初稿, 然后由王炳武修改定稿.

译者感谢渤海大学张守波副校长, 研究生院姜德刚院长, 数理学院王志福院长、张盛副院长对本书翻译工作的支持. 本书的翻译也得到了原著作者 Mordukhovich 教授和 Nam 教授的支持和鼓励, 在此表示感谢. 感谢第一译者的硕士研究生王超、鲁红、沈璐、韩冬雪、王凤娇、孙圆圆用 CTEX 录入了中译本的原稿. 中译本的出版得到了国家自然科学基金 (基金项目编号 11371070) 和国家重点基础研究发展计划课题项目 (项目编号 2011CB808902) 的资助, 在此表示感谢.

由于译者水平有限, 书中疏漏或不当之处在所难免, 恳请读者批评指正.

译　者

2015 年 7 月 31 日

前　　言

凸集和凸函数的一些几何性质在 20 世纪 60 年代之前已被许多杰出的数学家所研究 (其中对凸函数的研究程度相对低一些), 其中首推 Hermann Minkowski 和 Werner Fenchel. 20 世纪 60 年代初, R.Tyrrell Rockafellar 和 Jean-Jacques Moreau 的工作使得凸分析得到了很大的发展, 他们奠基了这个新领域的系统研究. 凸分析是变分分析的基础部分, 其中广义微分理论可用于研究初始数据不加任何可微假设的大量数学模型. 凸分析在许多应用中的重要性至今已经得到了公认, 这些应用领域中首先包括凸最优化. 凸性的存在不仅使全面地定性研究最优解以及导出最优性的必要和充分条件成为可能, 也有助于发展解凸最优化问题的行之有效的数值算法, 即使对不可微数据也如此. 凸分析和最优化理论正在对数学的许多方面及其应用发挥着日益增长的影响, 这些应用特别包括控制系统、评估与信号处理、通信与网络、电路设计、数据分析与建模、统计、经济与金融等.

现在已经有致力于凸分析与最优化的不同方面的基础书籍, 这里我们特别指出 Rockafellar 的《凸分析》[26], Hiriart-Urruty 与 Lemaréchal 的《凸分析与最小化算法》(两卷)[8] 及其精简版本[9], Borwein 与 Lewis 的《凸分析与非线性最优化》[4], Nesterov 的《凸最优化入门讲义》[21] 和 Boyd 与 Vandenberghe 的《凸最优化》[3], 以及书末参考文献中的其他书籍.

在凸分析与最优化这个大的框架下, 对刚刚开始利用凸分析接触该领域更深的课题的学生和研究人员来说, 本书可以充当一个桥梁. 书中陈述的大多结果都给出了详细的证明, 并包括了许多图表和练习, 以便更好地理解这些素材. 本书采用了现代变分分析中建立的强有力的几何方法, 这种方法在凸的情形得到了简化, 以此给读者提供了一个进入有限维空间中凸结构的广义微分理论的捷径, 因此本书也可作为感兴趣的读者继续研究非凸变分分析与应用的一个起点. 从此角度来说, 它对凸分析和变分分析方面的专家也可能是有意义的. 最后, 本书的应用部分不仅涉及了与最优性条件和次梯度算法有关的凸最优化的经典课题, 而且给出了一些近期的一类重要的选址问题的易于理解的定性和数值结果.

本书包括四章, 安排如下: 在第 1 章中我们研究凸集和凸函数的基本性质, 其中特别关注在最优化中起着重要作用的凸函数类. 第 2 章主要致力于建立凸集的法锥和凸函数次梯度的基本运算法则, 它们是凸理论的主流. 第 3 章包含了凸分析的其他一些课题, 它们在应用中有广泛的用途. 第 4 章则完全把注意力放在应用问题上, 从定性和数值两个角度讨论了凸分析的基本结果在凸最优化问题和选址问题

的某些选题中的应用. 最后, 书末给出了某些练习的答案和提示.

每章的最后都有练习题, 而图表和例子则贯穿全书. 参考文献包含一些书籍和选取的论文, 它们密切关联于本书所讨论的课题, 也可以帮助读者进一步学习更深的凸分析理论及其应用与扩展.

本书只需要线性代数和基础微积分的基本知识, 因此可用作本科生和研究生层次的凸分析及其应用课程的教科书. 事实上, 作者已经应用这些讲义在他们的大学以及作为访问学者在其他一些学校讲授过这样的课程. 我们希望本书能使本科生和研究生、不同学科的研究人员以及从业者等广泛群体更易于理解和进入凸分析这个领域.

Boris S. Mordukhovich, Nguyen Mau Nam

2013 年 12 月

符　号　表

$\mathbb{R}$	the real numbers (实数)
$\overline{\mathbb{R}} = (-\infty, \infty]$	the extended real line (增广实直线)
$\mathbb{R}_+$	the nonnegative real numbers (非负实数)
$\mathbb{R}_>$	the positive real numbers (正实数)
$\mathbb{N}$	the positive integers (正整数)
$\operatorname{co}\Omega$	convex hull of Ω (Ω 的凸包)
$\operatorname{aff}\Omega$	affine hull of Ω (Ω 的仿射包)
$\operatorname{int}\Omega$	interior of Ω (Ω 的内部)
$\operatorname{ri}\Omega$	relative interior of Ω (Ω 的相对内部)
$\operatorname{span}\Omega$	linear subspace generated by Ω (由 Ω 生成的线性子空间)
$\operatorname{cone}\Omega$	cone generated by Ω (由 Ω 生成的锥)
K_Ω	convex cone generated by Ω (由 Ω 生成的凸锥)
$\dim\Omega$	dimension of Ω (Ω 的维数)
$\overline{\Omega}$	closure of Ω (Ω 的闭包)
$\operatorname{bd}\Omega$	boundary of Ω (Ω 的边界)
$\mathbb{B}$	closed unit ball (闭单位球)
$\mathbb{B}(\bar{x}; r)$	closed ball with center $\bar{x}$ and radius r (以 $\bar{x}$ 为中心 r 为半径的闭球)
$\operatorname{dom} f$	domain of f (f 的定义域)
$\operatorname{epi} f$	epigraph of f (f 的上图)
$\operatorname{gph} F$	graph of mapping F (映射 F 的图)
$\mathcal{L}[a; b]$	line connecting a and b (连接 a 与 b 的直线)
$d(x; \Omega)$	distance from x to Ω (x 到 Ω 的距离)
$\Pi(x; \Omega)$	projection of x to Ω (x 到 Ω 的投影)
$\langle x, y\rangle$	inner product of x and y (x 与 y 的内积)
A^*	adjoint/transpose of linear mapping/matrix A (线性映射/矩阵 A 的共轭/转置)
$N(\bar{x}; \Omega)$	normal cone to Ω at $\bar{x}$ (Ω 在 $\bar{x}$ 的法锥)
$\ker A$	kernel of linear mapping A (线性映射 A 的核)
$D^*F(\bar{x}, \bar{y})$	coderivative to F at $(\bar{x}, \bar{y})$ (F 在 $(\bar{x}, \bar{y})$ 点的上导数)

$T(\bar{x};\Omega)$	tangent cone to Ω at $\bar{x}$ (Ω 在 $\bar{x}$ 的切锥)
F_∞	horizon/asymptotic cone of F (F 的地平/渐近锥)
$\mathcal{T}_\Omega^F$	minimal time function defined by constant dynamic F and target set Ω (由恒定动态 F 和目标集合 Ω 定义的极小时间函数)
$L(x,\lambda)$	Lagrangian (Lagrange 函数)
ρ_F	Minkowski gauge of F (F 的 Minkowski 度规)
$\Pi_F(\cdot;\Omega)$	generalized projection defined by the minimal time function (由极小时间函数定义的广义投影)
$G_F(\bar{x};t)$	generalized ball defined by dynamic F with center $\bar{x}$ and radius t (由以 $\bar{x}$ 为中心 t 为半径的动态 F 定义的广义球)

目　　录

第 1 章　凸集和凸函数

本章介绍 Euclid 空间 $\mathbb{R}^n$ 中凸集和凸函数的定义、例子和基本性质, 也包含了一些相关的材料.

1.1　预 备 知 识

我们从回顾 Euclid 空间 $\mathbb{R}^n$ 的经典概念和性质开始. 本节给出结果的证明可在高等微积分和线性代数的标准教科书中找到.

用 $\mathbb{R}^n$ 表示实数的 n 元数组 $x=(x_1,\cdots,x_n)$ 的全体构成的集合, 则在如下运算下 $\mathbb{R}^n$ 构成一个线性空间:

$$(x_1,\cdots,x_n)+(y_1,\cdots,y_n)=(x_1+y_1,\cdots,x_n+y_n),$$

$$\lambda(x_1,\cdots,x_n)=(\lambda x_1,\cdots,\lambda x_n),$$

其中 $(x_1,\cdots,x_n),(y_1,\cdots,y_n)\in\mathbb{R}^n,\lambda\in\mathbb{R}$. 在不致引起混淆的情况下, $\mathbb{R}^n$ 中的零元和 $\mathbb{R}$ 中的数零通常用相同的记号 0 表示.

对任意的 $x=(x_1,\cdots,x_n)\in\mathbb{R}^n$, 我们将其等同于列向量 $x=(x_1,\cdots,x_n)^{\mathrm{T}}$, 其中记号“T”表示向量的转置. 给定 $x=(x_1,\cdots,x_n)\in\mathbb{R}^n, y=(y_1,\cdots,y_n)\in\mathbb{R}^n$, x 与 y 的内积定义为

$$\langle x,y\rangle:=\sum_{i=1}^n x_iy_i.$$

下面的命题列出了 $\mathbb{R}^n$ 中内积的一些性质.

命题 1.1　对 $x,y,z\in\mathbb{R}^n,\lambda\in\mathbb{R}$, 有

(i) $\langle x,x\rangle\geqslant 0$, $\langle x,x\rangle=0$ 当且仅当 $x=0$.

(ii) $\langle x,y\rangle=\langle y,x\rangle$.

(iii) $\langle\lambda x,y\rangle=\lambda\langle x,y\rangle$.

(iv) $\langle x,y+z\rangle=\langle x,y\rangle+\langle x,z\rangle$.

$x=(x_1,\cdots,x_n)\in\mathbb{R}^n$ 的 Euclid 范数定义为

$$\|x\|:=\sqrt{x_1^2+\cdots+x_n^2}.$$

由定义直接得 $\|x\|=\sqrt{\langle x,x\rangle}$.

命题 1.2　对任意的 $x, y \in \mathbb{R}^n, \lambda \in \mathbb{R}$, 有

(i) $\|x\| \geqslant 0$, $\|x\| = 0$ 当且仅当 $x = 0$.

(ii) $\|\lambda x\| = |\lambda| \cdot \|x\|$.

(iii) $\|x + y\| \leqslant \|x\| + \|y\|$ (三角不等式).

(iv) $|\langle x, y\rangle| \leqslant \|x\| \cdot \|y\|$ (Cauchy-Schwarz 不等式).

利用 Euclid 范数可引入 $\mathbb{R}^n$ 中的球, 另外球可用于定义 $\mathbb{R}^n$ 中其他的拓扑概念.

定义 1.3　以 $\bar{x}$ 为中心, $r \geqslant 0$ 为半径的闭球和 $\mathbb{R}^n$ 的闭单位球分别定义为

$$\mathbb{B}(\bar{x}; r) := \{x \in \mathbb{R}^n \mid \|x - \bar{x}\| \leqslant r\}, \quad \mathbb{B} := \{x \in \mathbb{R}^n \mid \|x\| \leqslant 1\}.$$

易见, $\mathbb{B} = \mathbb{B}(0; 1), \mathbb{B}(\bar{x}; r) = \bar{x} + r\mathbb{B}$.

定义 1.4　设 $\Omega \subset \mathbb{R}^n$, 则 $\bar{x}$ 是 Ω 的一个内点, 如果存在 $\delta > 0$ 使得

$$\mathbb{B}(\bar{x}; \delta) \subset \Omega.$$

Ω 的所有内点构成的集合记为 $\operatorname{int}\Omega$. Ω 称为是开的, 如果 Ω 的每个点都是它的内点.

Ω 是开的当且仅当对每一个 $\bar{x} \in \Omega$, 存在 $\delta > 0$ 使得 $\mathbb{B}(\bar{x}; \delta) \subset \Omega$. 显然, 空集 $\varnothing$ 和全空间 $\mathbb{R}^n$ 是开的. 而且, 以 $\bar{x}$ 为中心 r 为半径的任意开球 $\mathbb{B}(\bar{x}; r) := \{x \in \mathbb{R}^n \mid \|x - \bar{x}\| < r\}$ 是开的.

定义 1.5　集合 $\Omega \subset \mathbb{R}^n$ 是闭的, 如果它的余集 $\Omega^c = \mathbb{R}^n \backslash \Omega$ 在 $\mathbb{R}^n$ 中是开的. 因此, 空集、全空间、任意球 $\mathbb{B}(\bar{x}; r)$ 在 $\mathbb{R}^n$ 中是闭的.

命题 1.6　(i) $\mathbb{R}^n$ 中任意开集族的并是开的.

(ii) $\mathbb{R}^n$ 中任意有限开集族的交是开的.

(iii) $\mathbb{R}^n$ 中任意闭集族的交是闭的.

(iv) $\mathbb{R}^n$ 中任意有限闭集族的并是闭的.

定义 1.7　设 $\{x_k\}$ 是 $\mathbb{R}^n$ 中的序列, 称 $\{x_k\}$ 收敛于 $\bar{x}$, 如果 $\|x_k - \bar{x}\| \to 0$(当 $k \to \infty$). 此时, 记

$$\lim_{k \to \infty} x_k = \bar{x}.$$

由定义 1.7 可用来定义下面关于集合的重要的拓扑概念.

定义 1.8　设 Ω 是 $\mathbb{R}^n$ 的非空子集, 则:

(i) Ω 的闭包, 记为 $\overline{\Omega}$ 或 $\operatorname{cl}\Omega$, 定义为属于 Ω 的所有收敛序列的极限的全体.

(ii) Ω 的边界, 记为 $\operatorname{bd}\Omega$, 定义为集合 $\overline{\Omega} \setminus \operatorname{int}\Omega$.

我们能看到, Ω 的闭包是包含 Ω 的所有闭集的交, Ω 的内部是包含在 Ω 中的所有开集的并. 由定义得 $\bar{x} \in \overline{\Omega}$ 当且仅当对任意的 $\delta > 0$, 有 $\mathbb{B}(\bar{x}; \delta) \cap \Omega \neq \varnothing$. 而且, $\bar{x} \in \operatorname{bd}\Omega$ 当且仅当对任意的 $\delta > 0$, 闭球 $\mathbb{B}(\bar{x}; \delta)$ 均与集合 Ω 和它余集 Ω^c 相交.

定义 1.9 设 $\{x_k\}$ 是 $\mathbb{R}^n$ 中的序列, $\{k_\ell\}$ 是严格递增的正整数数列, 则新序列 $\{x_{k_\ell}\}$ 称为 $\{x_k\}$ 的子序列.

称一个集合 Ω 是有界的, 如果它包含在一个以原点为中心, 以某 $r>0$ 为半径的球内, 即 $\Omega\subset\mathbb{B}(0;r)$. 于是, 序列 $\{x_k\}$ 是有界的, 如果存在 $r>0$ 使得

$$\|x_k\|\leqslant r(\forall k\in\mathbb{N}).$$

下面重要的结果称为 Bolzano-Weierstrass 定理.

定理 1.10 $\mathbb{R}^n$ 中的任何有界序列都包含收敛的子序列.

下面的概念在分析和最优化中都起着非常重要的作用.

定义 1.11 称集合 Ω 在 $\mathbb{R}^n$ 中是紧的, 如果 Ω 中的每个序列都有收敛于 Ω 中某点的子序列.

下面的结果是 Bolzano-Weierstrass 定理的推论.

定理 1.12 $\mathbb{R}^n$ 的子集 Ω 是紧的当且仅当它是闭的和有界的.

对于 $\mathbb{R}^n$ 的子集 Ω, Ω_1 和 Ω_2, 以及 $\lambda\in\mathbb{R}$, 定义运算:

$$\Omega_1+\Omega_2:=\{x+y\mid x\in\Omega_1,y\in\Omega_2\},\quad \lambda\Omega:=\{\lambda x\mid x\in\Omega\}.$$

易证下面的命题.

命题 1.13 设 Ω_1 和 Ω_2 是 $\mathbb{R}^n$ 的两个子集.

(i) 如果 Ω_1 是开的或者 Ω_2 是开的, 那么 $\Omega_1+\Omega_2$ 是开的.

(ii) 如果 Ω_1 是闭的且 Ω_2 是紧的, 那么 $\Omega_1+\Omega_2$ 是闭的.

先回顾实直线的子集有界性的概念.

定义 1.14 设 D 是实直线的子集, 数 $m\in\mathbb{R}$ 是 D 的一个下界, 如果有

$$x\geqslant m(\forall x\in D),$$

如果集合 D 有下界, 那么它是下有界的. 类似地, 数 $M\in\mathbb{R}$ 是 D 的一个上界, 如果

$$x\leqslant M(\forall x\in D),$$

那么 D 是上有界的, 如果它有一个上界. 而且, 称集合 D 是有界的, 如果它同时是下有界和上有界的.

下面给出集合的下确界和上确界的概念.

定义 1.15 设 $D\subset\mathbb{R}$ 是非空下有界的, D 的下确界, 记为 $\inf D$, 是 D 的最大下界. 当 D 是非空上有界时, 它的上确界, 记为 $\sup D$, 是 D 的最小上界. 如果 D 不是下有界 (上有界) 的, 则令 $\inf D:=-\infty(\sup D:=\infty)$. 我们约定 $\inf\varnothing:=\infty$, $\sup\varnothing:=-\infty$.

下面的基本公理确保这些概念是良好定义的.

完备性公理　对于 $\mathbb{R}$ 的任何非空子集 D, 如果它是上有界的, 那么它的最小上界存在且为一个实数.

利用完备性公理, 易见如果非空集合是下有界的, 那么它的最大下界存在且为一个实数.

全书为方便起见将考虑增广实值函数, 它在 $\overline{\mathbb{R}} := (-\infty, \infty]$ 中取值. 这里约定 $a + \infty = \infty (\forall a \in \mathbb{R})$, $\infty + \infty = \infty$, $t \cdot \infty = \infty (t > 0)$.

定义 1.16　设 $f : \Omega \to \overline{\mathbb{R}}$ 是增广实值函数, $\bar{x} \in \Omega$ 且 $f(\bar{x}) < \infty$, 则称 f 在 $\bar{x}$ 是连续的, 如果对 $\forall \varepsilon > 0$, $\exists \delta > 0$ 使得

$$|f(x) - f(\bar{x})| < \varepsilon \ (\|x - \bar{x}\| < \delta, \forall x \in \Omega).$$

称 f 在 Ω 上连续, 如果它在 Ω 中的任意点处是连续的.

显然, 由定义得如果 $f : \Omega \to \overline{\mathbb{R}}$ 在 $\bar{x}$ 是连续的(其中 $f(\bar{x}) < \infty$), 那么它在 Ω 和以 $\bar{x}$ 为中心, 以某 $r > 0$ 为半径的球的交集上是有限的. 而且, $f : \Omega \to \overline{\mathbb{R}}$ 在 $\bar{x}$ 连续 (其中 $f(\bar{x}) < \infty$) 当且仅当对 Ω 中收敛于 $\bar{x}$ 的任何序列 $\{x_k\}$, 有序列 $\{f(x_k)\}$ 收敛于 $f(\bar{x})$.

定义 1.17　设 $f : \Omega \to \overline{\mathbb{R}}$, $\bar{x} \in \Omega$ 且 $f(\bar{x}) < \infty$. 称 f 在 $\bar{x}$ 相对于 Ω 有局部极小值, 如果存在 $\delta > 0$, 使得

$$f(x) \geqslant f(\bar{x}), \quad \forall x \in \mathbb{B}(\bar{x}; \delta) \cap \Omega.$$

又称 f 在 $\bar{x}$ 相对于 Ω 有全局/绝对极小值, 如果

$$f(x) \geqslant f(\bar{x}), \quad \forall x \in \Omega.$$

局部和全局极大值的概念可类似地定义.

在本节的最后, 给出数学分析和最优化中被称为 Weierstrass 存在性定理的基本结论.

定理 1.18　设 $f : \Omega \to \mathbb{R}$ 是连续函数, 其中 Ω 是 $\mathbb{R}^n$ 的非空紧子集, 则存在 $\bar{x} \in \Omega$ 和 $\bar{u} \in \Omega$ 满足

$$f(\bar{x}) = \inf\{f(x) \mid x \in \Omega\}, \quad f(\bar{u}) = \sup\{f(x) \mid x \in \Omega\}.$$

在 4.1 节将给出一些定理 1.18 的“单边”版本.

1.2 凸　集

下面我们先研究集合凸性, 接着研究凸函数. 几何的思想在凸分析及其推广和应用中起着基本的作用, 因此在本书中将利用几何的方法.

给定 $\mathbb{R}^n$ 中的两个元素 a 和 b, 定义区间/线段

$$[a,b] := \{\lambda a + (1-\lambda)b \mid \lambda \in [0,1]\}^{①}.$$

注意到如果 $a=b$, 则区间简化为单点集 $[a,b]=\{a\}$.

定义 1.19 称 $\mathbb{R}^n$ 的子集 Ω 是凸的, 如果 $[a,b]\subset\Omega, \forall a,b\in\Omega$. 等价地, Ω 是凸的, 如果 $\lambda a+(1-\lambda)b\in\Omega, \forall a,b\in\Omega, \lambda\in[0,1]$ (图 1.1).

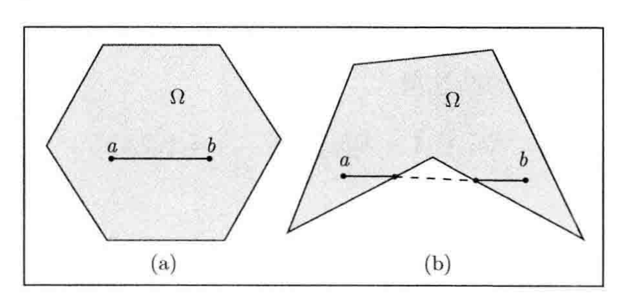


图 1.1 凸集和非凸集

给定 $m\in\mathbb{N}, \omega_1,\cdots,\omega_m\in\mathbb{R}^n$, 元素 $x=\sum_{i=1}^{m}\lambda_i\omega_i$ $\left(\text{其中 } \sum_{i=1}^{m}\lambda_i=1, \lambda_i\geqslant 0\right)$ 称为 $\omega_1,\cdots,\omega_m$ 的一个凸组合.

命题 1.20 $\mathbb{R}^n$ 的子集 Ω 是凸的当且仅当它包含它的元素的所有凸组合.

证明 充分性是平凡的. 为证必要性, 下面用数学归纳法证明 Ω 中元素的任意凸组合 $x=\sum_{i=1}^{m}\lambda_i\omega_i$ 是 Ω 中的元素. 当 $m=1,2$ 时, 根据定义直接得结论成立. 现固定正整数 $m\geqslant 2$, 假设 Ω 中任意 $k\in\mathbb{N}$ 个元素的凸组合 (其中 $k\leqslant m$) 属于 Ω. 给定凸组合

$$y := \sum_{i=1}^{m+1}\lambda_i\omega_i, \quad \sum_{i=1}^{m+1}\lambda_i = 1, \quad \lambda_i\geqslant 0.$$

注意到如果 $\lambda_{m+1}=1$, 则 $\lambda_1=\lambda_2=\cdots=\lambda_m=0$, 因此 $y=\omega_{m+1}\in\Omega$. 在 $\lambda_{m+1}<1$ 的情形, 有表达式

$$\sum_{i=1}^{m}\lambda_i = 1-\lambda_{m+1}, \quad \sum_{i=1}^{m}\frac{\lambda_i}{1-\lambda_{m+1}} = 1,$$

从而

$$z := \sum_{i=1}^{m}\frac{\lambda_i}{1-\lambda_{m+1}}\omega_i \in \Omega.$$

① 译者注: 通常的写法是 $(1-\lambda)a+\lambda b$, 如此则 λ 刚好是以 a 与 $(1-\lambda)a+\lambda b$ 为端点的线段的长度与线段 $[a,b]$ 长度的比, 且 λ 的取值是从 0 到 1, 刚好对应 $(1-\lambda)a+\lambda b$ 从 a 到 b. 原著写法的对应方向则相反.

因此有关系式

$$y=(1-\lambda_{m+1})\sum_{i=1}^{m}\frac{\lambda_i}{1-\lambda_{m+1}}\omega_i+\lambda_{m+1}\omega_{m+1}=(1-\lambda_{m+1})z+\lambda_{m+1}\omega_{m+1}\in\Omega,$$

这样就完成了命题的证明. □

命题 1.21 设 Ω_1 是 $\mathbb{R}^n$ 的凸子集, Ω_2 是 $\mathbb{R}^p$ 的凸子集, 则笛卡儿积 $\Omega_1\times\Omega_2$ 是 $\mathbb{R}^n\times\mathbb{R}^p$ 中的凸子集.

证明 取定 $a=(a_1,a_2), b=(b_1,b_2)\in\Omega_1\times\Omega_2$, $\lambda\in[0,1]$, 则有 $a_1,b_1\in\Omega_1$, $a_2,b_2\in\Omega_2$. 由 Ω_1 和 Ω_2 的凸性得

$$\lambda a_i+(1-\lambda)b_i\in\Omega_i,\quad i=1,2,$$

从而蕴涵着

$$\lambda a+(1-\lambda)b=(\lambda a_1+(1-\lambda)b_1,\lambda a_2+(1-\lambda)b_2)\in\Omega_1\times\Omega_2.$$

因此笛卡儿积 $\Omega_1\times\Omega_2$ 是凸的. □

下面给出仿射映射的定义.

定义 1.22 映射 $B:\mathbb{R}^n\to\mathbb{R}^p$ 是仿射的, 如果存在线性映射 $A:\mathbb{R}^n\to\mathbb{R}^p$ 和元素 $b\in\mathbb{R}^p$ 使得 $B(x)=A(x)+b,\forall\, x\in\mathbb{R}^n$.

每个线性映射都是仿射的. 而且, $B:\mathbb{R}^n\to\mathbb{R}^p$ 是仿射的当且仅当

$$B((\lambda x)+(1-\lambda)y)=\lambda B(x)+(1-\lambda)B(y),\quad\forall\, x,y\in\mathbb{R}^n,\quad\forall\,\lambda\in\mathbb{R}.$$

现证集合的凸性在仿射运算下保持不变.

命题 1.23 设 $B:\mathbb{R}^n\to\mathbb{R}^p$ 是仿射映射. 假设 Ω 是 $\mathbb{R}^n$ 的凸子集, Θ 是 $\mathbb{R}^p$ 的凸子集, 则 $B(\Omega)$ 是 $\mathbb{R}^p$ 的凸子集, $B^{-1}(\Theta)$ 是 $\mathbb{R}^n$ 的凸子集.

证明 任取 $a,b\in B(\Omega)$, $\lambda\in(0,1)$, 则有 $x,y\in\Omega$, 使得 $a=B(x)$, $b=B(y)$. 由于 Ω 是凸的, 有 $\lambda x+(1-\lambda)y\in\Omega$. 于是

$$\lambda a+(1-\lambda)b=\lambda B(x)+(1-\lambda)B(y)=B(\lambda x+(1-\lambda)y)\in B(\Omega),$$

这就证明了像 $B(\Omega)$ 的凸性.

现任取 $x,y\in B^{-1}(\Theta)$, $\lambda\in(0,1)$, 有 $B(x),B(y)\in\Theta$, 且由 Θ 的凸性有

$$\lambda B(x)+(1-\lambda)B(y)=B(\lambda x+(1-\lambda)y)\in\Theta.$$

因此, 有 $\lambda x+(1-\lambda)y\in B^{-1}(\Theta)$, 这就证明了逆像 $B^{-1}(\Theta)$ 的凸性. □

命题 1.24 设 $\Omega_1,\Omega_2\subset\mathbb{R}^n$ 是凸的, $\lambda\in\mathbb{R}$, 则 $\Omega_1+\Omega_2$, $\lambda\Omega_1$ 在 $\mathbb{R}^n$ 中也是凸的.

证明 由定义直接可得. □

接下来研究凸集的交的凸性.

命题 1.25　设 $\{\Omega_\alpha\}_{\alpha\in I}$ 是 $\mathbb{R}^n$ 的凸子集族, 则 $\bigcap_{\alpha\in I}\Omega_\alpha$ 也是 $\mathbb{R}^n$ 的凸子集.

证明　任取 $a,b\in\bigcap_{\alpha\in I}\Omega_\alpha$, 有 $a,b\in\Omega_\alpha$, $\forall\,\alpha\in I$. 每个 Ω_α 的凸性保证 $\lambda a+(1-\lambda)b\in\Omega_\alpha, \forall\,\lambda\in(0,1)$. 因此, $\lambda a+(1-\lambda)b\in\bigcap_{\alpha\in I}\Omega_\alpha$, 即交集 $\bigcap_{\alpha\in I}\Omega_\alpha$ 是凸的. □

定义 1.26　设 Ω 是 $\mathbb{R}^n$ 的子集, Ω 的凸包定义为

$$\operatorname{co}\Omega := \cap\{C \mid C\text{是凸的且}\Omega\subset C\}.$$

下面的命题由定义和命题 1.25 直接可得.

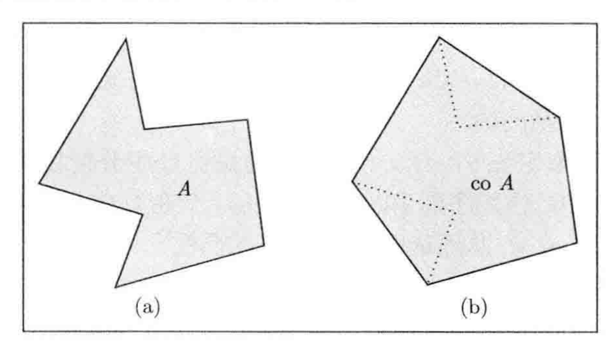


图 1.2　非凸集合和它的凸包

命题 1.27　凸包 $\operatorname{co}\Omega$ 是包含 Ω 的最小凸集.

证明　集合 $\operatorname{co}\Omega\supset\Omega$ 的凸性由命题 1.25 可得. 另外, 对包含 Ω 的任何凸集 C, 显然有 $\operatorname{co}\Omega\subset C$, 这就证明了结论成立. □

命题 1.28　对 $\mathbb{R}^n$ 中的任何子集 Ω, 它的凸包具有表达式

$$\operatorname{co}\Omega=\left\{\sum_{i=1}^m\lambda_i a_i\,\middle|\,\sum_{i=1}^m\lambda_i=1,\ \lambda_i\geqslant 0,\ a_i\in\Omega,\ m\in\mathbb{N}\right\}.$$

证明　用 C 表示表达式右边的集合, 则显然有 $\Omega\subset C$. 现验证集合 C 是凸的. 任取 $a,b\in C$, 则有

$$a:=\sum_{i=1}^p\alpha_i a_i,\quad b:=\sum_{j=1}^q\beta_j b_j,$$

其中 $a_i,b_j\in\Omega$, $\alpha_i,\beta_j\geqslant 0$ 且 $\sum_{i=1}^p\alpha_i=\sum_{j=1}^q\beta_j=1$, $p,q\in\mathbb{N}$. 对任意的数 $\lambda\in(0,1)$, 于是

$$\lambda a+(1-\lambda)b=\sum_{i=1}^p\lambda\alpha_i a_i+\sum_{j=1}^q(1-\lambda)\beta_j b_j.$$

由于

$$\sum_{i=1}^{p}\lambda\alpha_i+\sum_{j=1}^{q}(1-\lambda)\beta_j=\lambda\sum_{i=1}^{p}\alpha_i+(1-\lambda)\sum_{j=1}^{q}\beta_j=1,$$

所以 $\lambda a+(1-\lambda)b\in C$. 因此由 $\operatorname{co}\Omega$ 的定义得 $\operatorname{co}\Omega\subset C$. 现任取 $a=\sum_{i=1}^{m}\lambda_i a_i\in C$, 其中 $a_i\in\Omega\subset\operatorname{co}\Omega$, $i=1,\cdots,m$. 由于 $\operatorname{co}\Omega$ 是凸的, 由命题 1.20 得 $a\in\operatorname{co}\Omega$, 从而 $\operatorname{co}\Omega=C$. □

命题 1.29 凸集 $\Omega\subset\mathbb{R}^n$ 的内部 $\operatorname{int}\Omega$ 和闭包 $\overline{\Omega}$ 也是凸的.

证明 取定 $a,b\in\operatorname{int}\Omega$, $\lambda\in(0,1)$, 则找到开集 V 满足

$$a\in V\subset\Omega,$$

因此 $\lambda a+(1-\lambda)b\in\lambda V+(1-\lambda)b\subset\Omega$. 由于 $\lambda V+(1-\lambda)b$ 是开的, 得 $\lambda a+(1-\lambda)b\in\operatorname{int}\Omega$, 所以集合 $\operatorname{int}\Omega$ 是凸的.

为证 $\overline{\Omega}$ 的凸性, 取定 $a,b\in\overline{\Omega},\lambda\in(0,1)$, 则找到 Ω 中分别收敛于 a 与 b 的序列 $\{a_k\}$ 与 $\{b_k\}$. 由 Ω 的凸性知序列 $\{\lambda a_k+(1-\lambda)b_k\}$ 全在 Ω 中且收敛于 $\lambda a+(1-\lambda)b$. 这保证 $\lambda a+(1-\lambda)b\in\overline{\Omega}$, 从而证明了闭包 $\overline{\Omega}$ 的凸性. □

接下来, 对 $\forall a,b\in\mathbb{R}^n$, 定义区间

$$[a,b):=\{\lambda a+(1-\lambda)b\mid\lambda\in(0,1]\}.$$

用同样的方法也可定义区间 $(a,b]$ 和 (a,b).

引理 1.30 对具有非空内部的凸集 $\Omega\subset\mathbb{R}^n$, 任取 $a\in\operatorname{int}\Omega$, $b\in\overline{\Omega}$, 则 $[a,b)\subset\operatorname{int}\Omega$.

证明 由于 $b\in\overline{\Omega}$, 对 $\forall\varepsilon>0$, 有 $b\in\Omega+\varepsilon\mathbb{B}$. 现取 $\lambda\in(0,1]$, 令 $x_\lambda:=\lambda a+(1-\lambda)b$. 选取 $\varepsilon>0$ 使得 $a+\varepsilon\dfrac{2-\lambda}{\lambda}\mathbb{B}\subset\Omega$, 则得

$$\begin{aligned}x_\lambda+\varepsilon\mathbb{B}&=\lambda a+(1-\lambda)b+\varepsilon\mathbb{B}\\&\subset\lambda a+(1-\lambda)[\Omega+\varepsilon\mathbb{B}]+\varepsilon\mathbb{B}\\&=\lambda a+(1-\lambda)\Omega+(1-\lambda)\varepsilon\mathbb{B}+\varepsilon\mathbb{B}\\&\subset\lambda\left[a+\varepsilon\frac{2-\lambda}{\lambda}\mathbb{B}\right]+(1-\lambda)\Omega\\&\subset\lambda\Omega+(1-\lambda)\Omega\subset\Omega.\end{aligned}$$

这表明 $x_\lambda\in\operatorname{int}\Omega$, 从而证明了包含关系 $[a,b)\subset\operatorname{int}\Omega$. □

下面建立取凸集的内部和闭包之间的关系.

命题 1.31 设 $\Omega\subset\mathbb{R}^n$ 是具有非空内部的凸集, 则有

(i) $\overline{\operatorname{int}\Omega}=\overline{\Omega}$; (ii) $\operatorname{int}\Omega=\operatorname{int}\overline{\Omega}$.

证明 (i) 显然, $\overline{\mathrm{int}\Omega} \subset \overline{\Omega}$. 对 $\forall b \in \overline{\Omega}$, $a \in \mathrm{int}\,\Omega$, 定义序列 $\{x_k\}$ 如下

$$x_k := \frac{1}{k}a + \left(1 - \frac{1}{k}\right)b, \quad k \in \mathbb{N}.$$

引理 1.30 保证 $x_k \in \mathrm{int}\,\Omega$. 由于 $x_k \to b$, 有 $b \in \overline{\mathrm{int}\,\Omega}$, 因此证明了 (i) 成立.

(ii) 由于包含关系 $\mathrm{int}\,\Omega \subset \mathrm{int}\,\overline{\Omega}$ 是显然的, 下面证明相反的包含关系 $\mathrm{int}\,\overline{\Omega} \subset \mathrm{int}\,\Omega$. 为此, 对 $\forall b \in \mathrm{int}\,\overline{\Omega}$, $a \in \mathrm{int}\,\Omega$. 如果 $\varepsilon > 0$ 充分小, 则

$$c := b + \varepsilon(b - a) \in \overline{\Omega}, \quad \text{因此 } b = \frac{\varepsilon}{1+\varepsilon}a + \frac{1}{1+\varepsilon}c \in (a, c) \subset \mathrm{int}\,\Omega,$$

这就证明了 $\mathrm{int}\,\overline{\Omega} \subset \mathrm{int}\Omega$, 从而完成了证明. □

1.3 凸 函 数

本节收集了关于一般 (增广实值) 凸函数的基本事实, 包括它们的分析和几何刻画、重要的性质以及它们特殊子类的说明. 我们还定义了凸集值映射, 并且利用它们来研究下面用到的非常好的一类最优值函数.

定义 1.32 设 $f : \Omega \to \overline{\mathbb{R}}$ 是定义在凸集 $\Omega \subset \mathbb{R}^n$ 上的增广实值函数, 则函数 f 在 Ω 上是凸的, 如果

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y), \quad \forall x, y \in \Omega, \quad \lambda \in (0,1). \tag{1.1}$$

如果对 $x \neq y$, 式 (1.1) 中的不等式是严格的, 则 f 在 Ω 上是严格凸的.

给定函数 $f : \Omega \to \overline{\mathbb{R}}$, f 到 $\mathbb{R}^n$ 的延拓定义为

$$\tilde{f}(x) := \begin{cases} f(x), & x \in \Omega, \\ \infty, & \text{其他}. \end{cases}$$

显然, 如果 f 在 Ω 上是凸的, 这里 Ω 是凸集, 那么 $\tilde{f}$ 在 $\mathbb{R}^n$ 上是凸的. 而且, 如果 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸函数, 那么它在 $\mathbb{R}^n$ 上的任一凸子集上也是凸的. 不失一般性, 这允许在整个空间 $\mathbb{R}^n$ 上考虑增广实值凸函数.

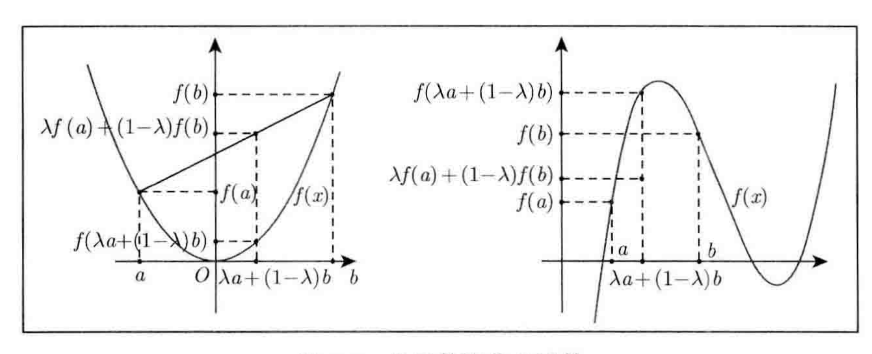


图 1.3 凸函数和非凸函数

定义 1.33 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 的定义域和上图分别定义为

$$\operatorname{dom} f := \{x\in\mathbb{R}^n \mid f(x)<\infty\}$$

和

$$\begin{aligned}\operatorname{epi} f :&= \{(x,t)\in\mathbb{R}^n\times\mathbb{R} \mid x\in\mathbb{R}^n,\, t\geqslant f(x)\}\\ &= \{(x,t)\in\mathbb{R}^n\times\mathbb{R} \mid x\in\operatorname{dom} f,\, t\geqslant f(x)\}.\end{aligned}$$

下面举例来说明函数的凸性.

例 1.34 下面的函数是凸的:

(i) $f(x):=\langle a,x\rangle+b, \forall\, x\in\mathbb{R}^n$, 这里 $a\in\mathbb{R}^n, b\in\mathbb{R}$.

(ii) $g(x):=\|x\|,\ \forall\, x\in\mathbb{R}^n$.

(iii) $h(x):=x^2,\ \forall\, x\in\mathbb{R}$.

事实上, (i) 中的函数 f 是凸的, 由于

$$\begin{aligned}f(\lambda x+(1-\lambda)y)&=\langle a,\lambda x+(1-\lambda)y\rangle+b=\lambda\langle a,x\rangle+(1-\lambda)\langle a,y\rangle+b\\ &=\lambda(\langle a,x\rangle+b)+(1-\lambda)(\langle a,y\rangle+b)\\ &=\lambda f(x)+(1-\lambda)f(y),\quad \forall\, x,y\in\mathbb{R}^n,\,\lambda\in(0,1).\end{aligned}$$

(ii) 中的函数 g 是凸的, 由于 $\forall\, x,y\in\mathbb{R}^n,\lambda\in(0,1)$, 有

$$g(\lambda x+(1-\lambda)y)=\|\lambda x+(1-\lambda)y\|\leqslant\lambda\|x\|+(1-\lambda)\|y\|=\lambda g(x)+(1-\lambda)g(y),$$

上式由三角不等式和事实 $\|\alpha u\|=|\alpha|\cdot\|u\|(\forall\,\alpha\in\mathbb{R},u\in\mathbb{R}^n)$ 直接可得.

(iii) 由下面的例子给出在 $\mathbb{R}^n$ 上二次函数中最简单的二次函数 h 的凸性的更一般结果.

例 1.35 设 A 是 $n\times n$ 对称矩阵, 它称为半正定的, 如果 $\langle Au,u\rangle\geqslant 0, \forall\, u\in\mathbb{R}^n$. 让我们验证 A 是半正定的当且仅当函数 $f:\mathbb{R}^n\to\mathbb{R}$ 定义为

$$f(x):=\frac{1}{2}\langle Ax,x\rangle,\quad x\in\mathbb{R}^n$$

是凸的. 事实上, 直接计算表明: 对 $\forall\, x,y\in\mathbb{R}^n,\lambda\in(0,1)$, 有

$$\lambda f(x)+(1-\lambda)f(y)-f(\lambda x+(1-\lambda)y)=\frac{1}{2}\lambda(1-\lambda)\langle A(x-y),x-y\rangle.\tag{1.2}$$

如果矩阵 A 是半正定的, 则 $\langle A(x-y),x-y\rangle\geqslant 0$, 因此根据式 (1.2) 函数 f 是凸的. 反之, 假设 f 的凸性并利用在等式 (1.2) 中取 $x=u,y=0$, 可证明 A 是半正定的.

定理 1.36 函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸的当且仅当对任意的数 $\lambda_i\geqslant 0(i=1,\cdots,m)$, 且 $\sum\limits_{i=1}^{m}\lambda_i=1$, 对任意的元 $x_i\in\mathbb{R}^n(i=1,\cdots,m)$, 下式成立:

$$f\left(\sum_{i=1}^{m}\lambda_i x_i\right)\leqslant\sum_{i=1}^{m}\lambda_i f(x_i).\tag{1.3}$$

证明 由于式 (1.3) 直接蕴涵着 f 的凸性, 只需证任何凸函数 f 满足 Jensen 不等式 (1.3). 利用数学归纳法证明, 考虑到当 $m=1$ 时不等式 (1.3) 平凡成立. 而且, 当 $m=2$ 时由凸性的定义知不等式 (1.3) 成立. 假设对整数 $m:=k(k\geqslant 2)$ 时不等式 (1.3) 成立. 取定数 $\lambda_i\geqslant 0, i=1,\cdots,k+1$, 且 $\sum_{i=1}^{k+1}\lambda_i=1$, 取定元素 $x_i\in\mathbb{R}^n, i=1,\cdots,k+1$. 显然有关系式

$$\sum_{i=1}^{k}\lambda_i=1-\lambda_{k+1}.$$

如果 $\lambda_{k+1}=1$, 则 $\lambda_i=0(i=1,\cdots,k)$ 且此时式 (1.3) 对 $m:=k+1$ 显然成立. 现假设 $0\leqslant\lambda_{k+1}<1$, 则

$$\sum_{i=1}^{k}\frac{\lambda_i}{1-\lambda_{k+1}}=1,$$

根据凸性和归纳假设直接计算得

$$\begin{aligned}
f\left(\sum_{i=1}^{k+1}\lambda_i x_i\right)&=f\left[(1-\lambda_{k+1})\frac{\sum_{i=1}^{k}\lambda_i x_i}{1-\lambda_{k+1}}+\lambda_{k+1}x_{k+1}\right]\\
&\leqslant(1-\lambda_{k+1})f\left(\frac{\sum_{i=1}^{k}\lambda_i x_i}{1-\lambda_{k+1}}\right)+\lambda_{k+1}f(x_{k+1})\\
&=(1-\lambda_{k+1})f\left(\sum_{i=1}^{k}\frac{\lambda_i}{1-\lambda_{k+1}}x_i\right)+\lambda_{k+1}f(x_{k+1})\\
&\leqslant(1-\lambda_{k+1})\sum_{i=1}^{k}\frac{\lambda_i}{1-\lambda_{k+1}}f(x_i)+\lambda_{k+1}f(x_{k+1})\\
&=\sum_{i=1}^{k+1}\lambda_i f(x_i),
\end{aligned}$$

这就证明了不等式 (1.3) 对 $m=k+1$ 成立, 从而完成了定理的证明. □

下面的定理通过相关的上图集合的凸性给出了函数凸性的几何刻画.

定理 1.37 函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸的当且仅当它的上图 $\operatorname{epi} f$ 是乘积空间 $\mathbb{R}^n\times\mathbb{R}$ 的凸子集.

证明　假设 f 是凸的, 取定 $(x_1,t_1),(x_2,t_2)\in \operatorname{epi} f$, $\lambda\in(0,1)$, 则有 $f(x_i)\leqslant t_i(i=1,2)$. 因此由 f 的凸性得

$$f(\lambda x_1+(1-\lambda)x_2)\leqslant \lambda f(x_1)+(1-\lambda)f(x_2)\leqslant \lambda t_1+(1-\lambda)t_2.$$

这蕴涵着

$$\lambda(x_1,t_1)+(1-\lambda)(x_2,t_2)=(\lambda x_1+(1-\lambda)x_2,\ \lambda t_1+(1-\lambda)t_2)\in \operatorname{epi} f,$$

从而上图 $\operatorname{epi} f$ 是 $\mathbb{R}^n\times\mathbb{R}$ 中的凸子集.

反之, 假设集合 $\operatorname{epi} f$ 是凸的, 取定 $x_1,x_2\in \operatorname{dom} f$, 数 $\lambda\in(0,1)$, 则 $(x_1,f(x_1))$, $(x_2,f(x_2))\in \operatorname{epi} f$. 因此, 有

$$(\lambda x_1+(1-\lambda)x_2,\lambda f(x_1)+(1-\lambda)f(x_2))=\lambda(x_1,f(x_1))+(1-\lambda)(x_2,f(x_2))\in \operatorname{epi} f,$$

从而

$$f(\lambda x_1+(1-\lambda)x_2)\leqslant \lambda f(x_1)+(1-\lambda)f(x_2)$$

成立, 这就证明了函数 f 的凸性. □

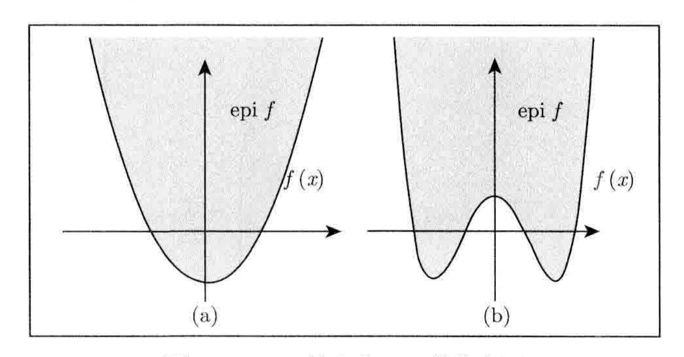


图 1.4　凸函数和非凸函数的上图

现在, 证明凸性保持在某些重要的运算中.

命题 1.38　设 $f_i:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸函数, $i=1,\cdots,m$, 则下面的函数也是凸函数:

(i) 对任意的 $\lambda>0$, 标量乘积 λf.

(ii) 和函数 $\displaystyle\sum_{i=1}^m f_i$.

(iii) 最大值函数 $\max_{1\leqslant i\leqslant m} f_i$.

证明　(i) 中 λf 的凸性由定义直接可得. 由于一般的情形由数学归纳法立即可得, 只需证明当 $m=2$ 时 (ii) 与 (iii) 中函数的凸性.

(ii) 任取 $x, y \in \mathbb{R}^n, \lambda \in (0,1)$, 则有

$$\begin{aligned}(f_1+f_2)(\lambda x+(1-\lambda)y) &= f_1(\lambda x+(1-\lambda)y)+f_2(\lambda x+(1-\lambda)y)\\ &\leqslant \lambda f_1(x)+(1-\lambda)f_1(y)+\lambda f_2(x)+(1-\lambda)f_2(y)\\ &= \lambda(f_1+f_2)(x)+(1-\lambda)(f_1+f_2)(y),\end{aligned}$$

这就证明了和函数 f_1+f_2 的凸性.

(iii) 记 $g := \max\{f_1, f_2\}$, 任取 $x, y \in \mathbb{R}^n, \lambda \in (0,1)$, 对 $i=1,2$, 得

$$f_i(\lambda x+(1-\lambda)y) \leqslant \lambda f_i(x)+(1-\lambda)f_i(y) \leqslant \lambda g(x)+(1-\lambda)g(y).$$

所以

$$\begin{aligned}g(\lambda x+(1-\lambda)y) &= \max\{f_1(\lambda x+(1-\lambda)y), f_2(\lambda x+(1-\lambda)y)\}\\ &\leqslant \lambda g(x)+(1-\lambda)g(y),\end{aligned}$$

这就证明了最大值函数 $g(x)=\max\{f_1(x), f_2(x)\}$ 是凸的. □

命题 1.39 设 $f:\mathbb{R}^n \to \mathbb{R}$ 是凸的, $\varphi:\mathbb{R}\to\overline{\mathbb{R}}$ 在包含函数 f 的值域的凸集上是非减的和凸的, 则复合 $\varphi\circ f$ 是凸的.

证明 任取 $x_1, x_2 \in \mathbb{R}^n, \lambda \in (0,1)$. 由 f 的凸性得

$$f(\lambda x_1+(1-\lambda)x_2) \leqslant \lambda f(x_1)+(1-\lambda)f(x_2).$$

由于 φ 是非减的, 且是凸的, 所以有

$$\begin{aligned}(\varphi\circ f)(\lambda x_1+(1-\lambda)x_2) &= \varphi(f(\lambda x_1+(1-\lambda)x_2))\\ &\leqslant \varphi(\lambda f(x_1)+(1-\lambda)f(x_2))\\ &\leqslant \lambda\varphi(f(x_1))+(1-\lambda)\varphi(f(x_2))\\ &= \lambda(\varphi\circ f)(x_1)+(1-\lambda)(\varphi\circ f)(x_2),\end{aligned}$$

这就证明了复合 $\varphi\circ f$ 的凸性. □

现在考虑凸函数和仿射映射的复合.

命题 1.40 设 $B:\mathbb{R}^n\to\mathbb{R}^p$ 是仿射映射, $f:\mathbb{R}^p\to\overline{\mathbb{R}}$ 是凸函数, 则复合 $f\circ B$ 是凸的.

证明 任取 $x, y \in \mathbb{R}^n$, $\lambda \in (0,1)$, 有

$$\begin{aligned}(f\circ B)(\lambda x+(1-\lambda)y) &= f(B(\lambda x+(1-\lambda)y)) = f(\lambda B(x)+(1-\lambda)B(y))\\ &\leqslant \lambda f(B(x))+(1-\lambda)f(B(y))\\ &= \lambda(f\circ B)(x)+(1-\lambda)(f\circ B)(y),\end{aligned}$$

这就证明了复合 $f \circ B$ 的凸性. □

下面关于命题 1.40 的简单推论在应用中是有用的.

推论 1.41 设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸函数. 对任意的 $\bar{x}, d \in \mathbb{R}^n$, 定义为 $\varphi(t) := f(\bar{x} + td)$ 的函数 $\varphi : \mathbb{R} \to \overline{\mathbb{R}}$ 也是凸的. 反之, 如果对任意的 $\bar{x}, d \in \mathbb{R}^n$, 如上定义的函数 φ 是凸的, 则 f 也是凸的.

证明 由于 $B(t) = \bar{x} + td$ 是仿射映射, φ 的凸性由命题 1.40 直接可得. 为证相反蕴涵关系, 任取 $x_1, x_2 \in \mathbb{R}^n, \lambda \in (0,1)$, 令 $\bar{x} := x_2, d := x_1 - x_2$. 由于 $\varphi(t) = f(\bar{x} + td)$ 是凸的, 有

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2) &= f(x_2 + \lambda(x_1 - x_2)) = \varphi(\lambda) = \varphi(\lambda(1) + (1-\lambda)(0)) \\ &\leqslant \lambda\varphi(1) + (1-\lambda)\varphi(0) = \lambda f(x_1) + (1-\lambda) f(x_2), \end{aligned}$$

这就证明了函数 f 的凸性. □

命题 1.42 设 $f : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$ 是凸的, 则对 $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^p$, 函数 $\varphi(y) := f(\bar{x}, y)$ 和 $\psi(x) := f(x, \bar{y})$ 也是凸的.

现在我们给出命题 1.38(iii) 的一个重要推广.

命题 1.43 设 $f_i : \mathbb{R}^n \to \overline{\mathbb{R}} (i \in I)$ 是具有非空指标集 I 的凸函数族, 则上确界函数 $f(x) := \sup_{i \in I} f_i(x)$ 是凸的.

证明 取定 $x_1, x_2 \in \mathbb{R}^n, \lambda \in (0,1)$. 对 $\forall i \in I$, 有

$$f_i(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f_i(x_1) + (1-\lambda) f_i(x_2) \leqslant \lambda f(x_1) + (1-\lambda) f(x_2),$$

上式相应地蕴涵着

$$f(\lambda x_1 + (1-\lambda)x_2) = \sup_{i \in I} f_i(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f(x_1) + (1-\lambda) f(x_2).$$

这就证明了上确界函数的凸性. □

接下来刻画单变量光滑函数的凸性. 为此, 从下面的引理开始.

引理 1.44 给定凸函数 $f : \mathbb{R} \to \overline{\mathbb{R}}$, 假设它的定义域是开区间 I. 对任意的 $a, b \in I$ 且 $a < x < b$, 有不等式

$$\frac{f(x) - f(a)}{x - a} \leqslant \frac{f(b) - f(a)}{b - a} \leqslant \frac{f(b) - f(x)}{b - x}.$$

证明 取定 a, b, x 如上, 构造数 $t := \dfrac{x-a}{b-a} \in (0,1)$, 则

$$f(x) = f(a + (x-a)) = f\left(a + \frac{x-a}{b-a}(b-a)\right) = f(a + t(b-a)) = f(tb + (1-t)a).$$

由上式可得不等式 $f(x) \leqslant tf(b) + (1-t)f(a)$ 和

$$f(x) - f(a) \leqslant tf(b) + (1-t)f(a) - f(a) = t[f(b) - f(a)] = \frac{x-a}{b-a}(f(b) - f(a)),$$

上式可等价地写为

$$\frac{f(x) - f(a)}{x-a} \leqslant \frac{f(b) - f(a)}{b-a}.$$

类似地, 有估计

$$\begin{aligned} f(x) - f(b) &\leqslant tf(b) + (1-t)f(a) - f(b) \\ &= (t-1)[f(b) - f(a)] = \frac{x-b}{b-a}(f(b) - f(a)), \end{aligned}$$

上式最后蕴涵着

$$\frac{f(b) - f(a)}{b-a} \leqslant \frac{f(b) - f(x)}{b-x},$$

从而完成了引理的证明. □

定理 1.45 假设 $f : \mathbb{R} \to \overline{\mathbb{R}}$ 在它的定义域上是可微的, 这里定义域是一开区间 I, 则 f 是凸的当且仅当导数 f' 在 I 上是非减的.

证明 假设 f 是凸的, 取定 $a < b, a, b \in I$, 根据引理 1.44, 有

$$\frac{f(x) - f(a)}{x-a} \leqslant \frac{f(b) - f(a)}{b-a}.$$

对任意的 $x \in (a, b)$ 成立. 由导数的定义, 上式蕴涵着

$$f'(a) \leqslant \frac{f(b) - f(a)}{b-a}.$$

类似地, 得估计

$$\frac{f(b) - f(a)}{b-a} \leqslant f'(b),$$

因此 $f'(a) \leqslant f'(b)$, 即 f' 是非减函数.

为证相反的结论成立, 假设 f' 在 I 上是非减的, 且取定 $x_1 < x_2, x_1, x_2 \in I, t \in (0, 1)$, 则

$$x_1 < x_t < x_2, \quad \text{其中} \quad x_t := tx_1 + (1-t)x_2.$$

利用中值定理, 可找到 c_1, c_2 使得 $x_1 < c_1 < x_t < c_2 < x_2$ 且

$$f(x_t) - f(x_1) = f'(c_1)(x_t - x_1) = f'(c_1)(1-t)(x_2 - x_1),$$

$$f(x_t) - f(x_2) = f'(c_2)(x_t - x_2) = f'(c_2)t(x_1 - x_2).$$

上式可等价地改写为

$$tf(x_t) - tf(x_1) = f'(c_1)t(1-t)(x_2 - x_1),$$

$$(1-t)f(x_t) - (1-t)f(x_2) = f'(c_2)t(1-t)(x_1 - x_2).$$

由于 $f'(c_1) \leqslant f'(c_2)$, 将以上两个等式相加得

$$f(x_t) \leqslant tf(x_1) + (1-t)f(x_2),$$

这就证明了函数 f 的凸性. □

推论 1.46　设 $f:\mathbb{R}\to\overline{\mathbb{R}}$ 在它的定义域上是二次可微的, 这里定义域是开区间 I, 则 f 是凸的当且仅当 $f''(x)\geqslant 0, \forall x\in I$.

证明　由于 $f''(x)\geqslant 0(\forall x\in I)$ 当且仅当导函数 f' 在该区间上是非减的, 则结论由定理 1.45 直接可得. □

例 1.47　考虑函数

$$f(x):=\begin{cases}\dfrac{1}{x}, & x>0,\\ \infty, & \text{其他}.\end{cases}$$

由于对所有属于 f 的定义域 (它是 $I=(0,\infty)$) 的 x, 有 $f''(x)=\dfrac{2}{x^3}>0$. 因此根据推论 1.46 这个函数在 $\mathbb{R}$ 上是凸的.

接下来定义集值映射的概念, 它在现代凸分析及其推广与应用中起着重要的作用.

定义 1.48　称 F 是 $\mathbb{R}^n$ 到 $\mathbb{R}^p$ 的集值映射, 记为 $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^p$, 如果对任意的 $x\in\mathbb{R}^n, F(x)$ 是 $\mathbb{R}^p$ 中的子集. F 的定义域和图分别定义为

$$\operatorname{dom}F:=\{x\in\mathbb{R}^n \mid F(x)\neq\varnothing\},\quad \operatorname{gph}F:=\{(x,y)\in\mathbb{R}^n\times\mathbb{R}^p \mid y\in F(x)\}.$$

任何单值映射 $F:\mathbb{R}^n\to\mathbb{R}^p$ 都是特殊的集值映射, 其中对任意的 $x\in\mathbb{R}^n$ 集合 $F(x)$ 是单点集. 在下面的定义中映射 F 的集值性是很关键的.

定义 1.49　设 $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^p, \varphi:\mathbb{R}^n\times\mathbb{R}^p\to\overline{\mathbb{R}}$. 与 F 和 φ 相关的最优值或边际函数定义为

$$\mu(x):=\inf\{\varphi(x,y) \mid y\in F(x)\},\quad x\in\mathbb{R}^n. \tag{1.4}$$

本节假设 $\mu(x)>-\infty(\forall x\in\mathbb{R}^n)$.

命题 1.50　假设 $\varphi:\mathbb{R}^n\times\mathbb{R}^p\to\overline{\mathbb{R}}$ 是凸函数, $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^p$ 是凸图的, 则式(1.4)中的最优值函数 μ 是凸的.

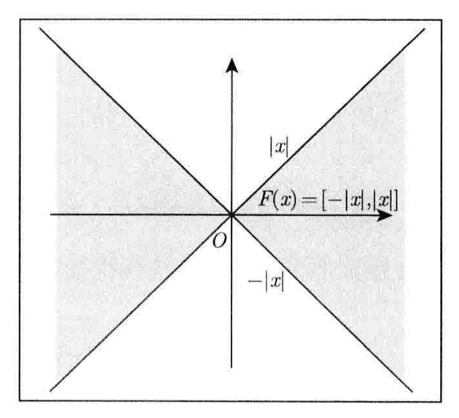


图 1.5 集值映射的图

证明 取 $x_1, x_2 \in \operatorname{dom}\mu, \lambda \in (0,1)$. 对 $\forall \varepsilon > 0$, 找到 $y_i \in F(x_i)$ 满足

$$\varphi(x_i, y_i) < \mu(x_i) + \varepsilon, \quad i = 1, 2.$$

它直接蕴涵着不等式

$$\lambda\varphi(x_1, y_1) < \lambda\mu(x_1) + \lambda\varepsilon, \quad (1-\lambda)\varphi(x_2, y_2) < (1-\lambda)\mu(x_2) + (1-\lambda)\varepsilon.$$

把这两个不等式相加, 并利用 φ 的凸性得

$$\begin{aligned}\varphi(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) &\leqslant \lambda\varphi(x_1, y_1) + (1-\lambda)\varphi(x_2, y_2) \\ &< \lambda\mu(x_1) + (1-\lambda)\mu(x_2) + \varepsilon.\end{aligned}$$

另外, 由 $\operatorname{gph} F$ 的凸性得

$$(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) = \lambda(x_1, y_1) + (1-\lambda)(x_2, y_2) \in \operatorname{gph} F,$$

因此 $\lambda y_1 + (1-\lambda)y_2 \in F(\lambda x_1 + (1-\lambda)x_2)$. 这蕴涵着

$$\begin{gathered}\mu(\lambda x_1 + (1-\lambda)x_2) \leqslant \varphi(\lambda x_1 + (1-\lambda)x_2, \\ \lambda y_1 + (1-\lambda)y_2) < \lambda\mu(x_1) + (1-\lambda)\mu(x_2) + \varepsilon.\end{gathered}$$

最后, 令 $\varepsilon \to 0$ 得最优值函数 μ 的凸性. □

利用命题 1.50, 可以证明许多情形下的凸性. 例如, 给定两个凸函数 $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, 2$, 令 $\varphi(x, y) := f_1(x) + y$, $F(x) := [f_2(x), \infty)$, 则函数 φ 是凸的, 且集合 $\operatorname{gph} F = \operatorname{epi} f_2$ 也是凸的, 因此证明了和

$$\mu(x) = \inf_{y \in F(x)} \varphi(x, y) = f_1(x) + f_2(x)$$

的凸性.

还有一个例子涉及复合. 设 $f:\mathbb{R}^p\to\overline{\mathbb{R}}$ 是凸的, $B:\mathbb{R}^n\to\mathbb{R}^p$ 是仿射的. 定义 $\varphi(x,y):=f(y),F(x):=\{B(x)\}$. 注意到 φ 是凸的, 而 F 是凸图的, 因此有凸复合

$$\mu(x)=\inf_{y\in F(x)}\varphi(x,y)=f(B(x)),\quad x\in\mathbb{R}^n.$$

这个例子重新给出了以前直接证明的结果. 现在利用命题 1.50 建立三类新函数的凸性.

命题 1.51　设 $\varphi:\mathbb{R}^p\to\overline{\mathbb{R}}$ 是凸的, $B:\mathbb{R}^p\to\mathbb{R}^n$ 是仿射的. 考虑集值逆像映射 $B^{-1}:\mathbb{R}^n\rightrightarrows\mathbb{R}^p$, 定义

$$f(x):=\inf\{\varphi(y)\mid y\in B^{-1}(x)\},\quad x\in\mathbb{R}^n,$$

假设对任意的 $x\in\mathbb{R}^n$, $f(x)>-\infty$, 则 f 是凸函数.

证明　令 $\varphi(x,y)\equiv\varphi(y)$, $F(x):=B^{-1}(x)$, 则集合

$$\operatorname{gph}F=\{(u,v)\in\mathbb{R}^n\times\mathbb{R}^p\mid B(v)=u\}$$

显然是凸的. 由于有表示

$$f(x)=\inf_{y\in F(x)}\varphi(y),\quad x\in\mathbb{R}^n,$$

所以 f 的凸性由命题 1.50 直接可得. □

命题 1.52　对凸函数 $f_1,f_2:\mathbb{R}^n\to\overline{\mathbb{R}}$, 定义下卷积:

$$(f_1\oplus f_2)(x):=\inf\{f_1(x_1)+f_2(x_2)\mid x_1+x_2=x\}.$$

假设对任意的 $x\in\mathbb{R}^n,(f_1\oplus f_2)(x)>-\infty$, 则 $f_1\oplus f_2$ 也是凸的.

证明　定义 $\varphi:\mathbb{R}^n\times\mathbb{R}^n\to\overline{\mathbb{R}}$ 为 $\varphi(x_1,x_2):=f_1(x_1)+f_2(x_2)$, 定义 $B:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}^n$ 为 $B(x_1,x_2):=x_1+x_2$, 则有

$$\inf\{\varphi(x_1,x_2)\mid(x_1,x_2)\in B^{-1}(x)\}=(f_1\oplus f_2)(x),\quad\forall x\in\mathbb{R}^n,$$

根据命题 1.51, 上式蕴涵着 $(f_1\oplus f_2)$ 的凸性. □

定义 1.53　函数 $g:\mathbb{R}^p\to\overline{\mathbb{R}}$ 称为分量非递减的, 如果

$$[x_i\leqslant y_i,i=1,\cdots,p]\Rightarrow[g(x_1,\cdots,x^p)\leqslant g(y_1,\cdots,y^p)].$$

现在将在本节给出命题 1.50 关于复合的最后一个结果.

命题 1.54 定义 $h:\mathbb{R}^n\to\mathbb{R}^p$ 为 $h(x):=(f_1(x),\cdots,f_p(x))$, 其中 $f_i:\mathbb{R}^n\to\mathbb{R}(i=1,\cdots,p)$ 是凸函数. 假设 $g:\mathbb{R}^p\to\overline{\mathbb{R}}$ 是凸的, 非分量递减的, 则复合 $g\circ h:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸函数.

证明 设 $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^p$ 是定义为

$$F(x):=[f_1(x),\infty)\times[f_2(x),\infty)\times\cdots\times[f_p(x),\infty)$$

的集值映射, 则 F 的图可表示为

$$\operatorname{gph}F=\{(x,t_1,\cdots,t_p)\in\mathbb{R}^n\times\mathbb{R}^p\mid t_i\geqslant f_i(x)\}.$$

由于所有的 f_i 是凸的, 所以集合 $\operatorname{gph}F$ 也是凸的. 进一步, 定义 $\varphi:\mathbb{R}^n\times\mathbb{R}^p\to\overline{\mathbb{R}}$ 为 $\varphi(x,y):=g(y)$, 并且注意到, 由于 g 是分量非递减的, 所以有

$$\inf\{\varphi(x,y)\mid y\in F(x)\}=g(f_1(x),\cdots,f_p(x))=(g\circ h)(x),$$

根据命题 1.50, 这就保证了复合 $g\circ h$ 的凸性. □

1.4 凸集的相对内部

本节首先介绍仿射集的定义和性质. 给定 $\mathbb{R}^n$ 中的两个元素 a 和 b, 连接它们的直线是

$$\mathcal{L}[a,b]:=\{\lambda a+(1-\lambda)b\mid\lambda\in\mathbb{R}\}.$$

注意到如果 $a=b$, 则 $\mathcal{L}[a,b]=\{a\}$.

定义 1.55 $\mathbb{R}^n$ 的子集 Ω 是仿射的, 如果对任意的 $a,b\in\Omega$, 有 $\mathcal{L}[a,b]\subset\Omega$.

例如, $\mathbb{R}^3$ 中的任意点、直线和平面是仿射集. 空集和全空间总是仿射的. 由定义可得任意仿射集族的交是仿射的. 这引出集合的仿射包的构造.

定义 1.56 集合 $\Omega\subset\mathbb{R}^n$ 的仿射包是

$$\operatorname{aff}\Omega:=\cap\{C\mid C\ \text{是仿射的且}\ \Omega\subset C\},$$

具有形式

$$x=\sum_{i=1}^m\lambda_i\omega_i,\quad\text{其中}\quad\sum_{i=1}^m\lambda_i=1,\quad m\in\mathbb{N}$$

的在 $\mathbb{R}^n$ 中的元素 x 称为 $\omega_1,\cdots,\omega_m$ 的仿射组合.

下一个命题的证明是直截了当的, 因此这里略去.

命题 1.57 以下结论成立:

(i) 集合 Ω 是仿射的当且仅当 Ω 包含它的元素的所有仿射组合.

(ii) 设 Ω, Ω_1 和 Ω_2 是 $\mathbb{R}^n$ 的仿射子集, 则和 $\Omega_1+\Omega_2$, 数量积 $\lambda\Omega(\forall\lambda\in\mathbb{R})$ 也是 $\mathbb{R}^n$ 的仿射子集.

(iii) 设 $B:\mathbb{R}^n\to\mathbb{R}^p$ 是仿射映射, 如果 Ω 是 $\mathbb{R}^n$ 的仿射子集, Θ 是 $\mathbb{R}^p$ 的仿射子集, 则像 $B(\Omega)$ 是 $\mathbb{R}^p$ 的仿射子集, 逆像 $B^{-1}(\Theta)$ 是 $\mathbb{R}^n$ 的仿射子集.

(iv) 给定 $\Omega\subset\mathbb{R}^n$, 它的仿射包是包含 Ω 的最小仿射集, 有

$$\operatorname{aff}\Omega=\left\{\sum_{i=1}^m\lambda_i\omega_i\,\middle|\,\sum_{i=1}^m\lambda_i=1,\omega_i\in\Omega,m\in\mathbb{N}\right\}.$$

(v) 集合 Ω 是 (线性) 子空间当且仅当 Ω 是包含原点的仿射集.

接下来考虑仿射集和 (线性) 子空间之间的关系.

引理 1.58　$\mathbb{R}^n$ 的非空子集 Ω 是仿射的当且仅当对任意的 $\omega\in\Omega$, $\Omega-\omega$ 是 $\mathbb{R}^n$ 的子空间.

证明　假设非空集合 $\Omega\subset\mathbb{R}^n$ 是仿射的. 由命题 1.57(v) 知对任意的 $\omega\in\Omega$, $\Omega-\omega$ 是子空间. 反之, 取定 $\omega\in\Omega$, 假设 $\Omega-\omega$ 是子空间记为 L, 则集合 $\Omega=\omega+L$ 显然是仿射的. □

引理 1.58 引出下面的概念.

定义 1.59　仿射集 Ω 是与子空间 L 平行的, 如果对某 $\omega\in\Omega$, $\Omega=\omega+L$.

下面的命题对平行子空间的形式作出解释.

命题 1.60　设 Ω 是 $\mathbb{R}^n$ 的非空仿射子集, 则它平行于 $\mathbb{R}^n$ 中的唯一子空间 L, 其中 $L=\Omega-\Omega$.

证明　给定非空仿射集 Ω, 取定 $\omega\in\Omega$, 得到线性子空间 $L:=\Omega-\omega$ 平行于 Ω. 为证明这样 L 的唯一性, 任取 $\omega_1,\omega_2\in\Omega$, 任取子空间 $L_1,L_2\subset\mathbb{R}^n$ 使得 $\Omega=\omega_1+L_1=\omega_2+L_2$, 则 $L_1=\omega_2-\omega_1+L_2$. 由于 $0\in L_1$, 有 $\omega_1-\omega_2\in L_2$. 这蕴涵着 $\omega_2-\omega_1\in L_2$, 从而 $L_1=\omega_2-\omega_1+L_2\subset L_2$. 类似地, 得 $L_2\subset L_1$, 这就证明了 $L_1=L_2$.

余下需证明表示 $L=\Omega-\Omega$. 设 $\Omega=\omega+L$, 其中 L 是与 Ω 平行的唯一子空间, $\omega\in\Omega$, 则 $L=\Omega-\omega\subset\Omega-\Omega$. 任取 $x=u-\omega$, 其中 $u,\omega\in\Omega$, 并且注意到 $\Omega-\omega$ 是平行于 Ω 的子空间. 因此根据上面证明的 L 的唯一性. $\Omega-\omega=L$. 这保证 $x\in\Omega-\omega=L$, 因此 $\Omega-\Omega=L$. □

平行子空间的唯一性表明下面的概念是有定义的.

定义 1.61　仿射集 $\varnothing\neq\Omega\subset\mathbb{R}^n$ 的维数定义为与 Ω 平行的线性子空间的维数. 而且, 凸集 $\varnothing\neq\Omega\subset\mathbb{R}^n$ 的维数是它的仿射包 $\operatorname{aff}\Omega$ 的维数.

接下来, 还需要另外一个定义, 这在下面是重要的.

定义 1.62 $\mathbb{R}^n$ 中的元素 $v_0, \cdots, v_m$, $m \geqslant 1$, 是仿射无关的, 如果

$$\left[\sum_{i=0}^{m} \lambda_i v_i = 0, \sum_{i=0}^{m} \lambda_i = 0\right] \Rightarrow [\lambda_i = 0, i = 0, \cdots, m].$$

易见与线性无关性有下面的关系.

命题 1.63 $\mathbb{R}^n$ 中的元素 $v_0, \cdots, v_m$ 是仿射无关的当且仅当元素 $v_1 - v_0, \cdots, v_m - v_0$ 是线性无关的.

证明 假设 $v_1 - v_0, \cdots, v_m - v_0$ 是仿射无关的, 考虑系统

$$\sum_{i=1}^{m} \lambda_i (v_i - v_0) = 0, \quad \text{即} \quad \lambda_0 v_0 + \sum_{i=1}^{m} \lambda_i v_i = 0,$$

其中 $\lambda_0 := -\sum_{i=1}^{m} \lambda_i$, 由于元素 $v_0, \cdots, v_m$ 是仿射无关的且 $\sum_{i=0}^{m} \lambda_i = 0$, 有 $\lambda_i = 0, i = 1, \cdots, m$, 所以 $v_1 - v_0, \cdots, v_m - v_0$ 是线性无关的. 相反的结论的证明是直截了当的. □

回顾某集合 C 的扩张, 即 $\operatorname{span} C$, 是指由 C 生成的线性子空间.

引理 1.64 设 $\Omega := \operatorname{aff}\{v_0, \cdots, v_m\}$, 其中 $v_i \in \mathbb{R}^n, i = 0, \cdots, m$, 则集合 $\{v_1 - v_0, \cdots, v_m - v_0\}$ 的扩张是平行于 Ω 的子空间.

证明 记 L 为平行于 Ω 的子空间, 则 $\Omega - v_0 = L$, 因此对所有的 $i = 1, \cdots, m$, $v_i - v_0 \in L$. 从而

$$\operatorname{span}\{v_i - v_0 \mid i = 1, \cdots, m\} \subset L.$$

为证相反的包含关系, 任取 $v \in L$, 有 $v + v_0 \in \Omega$. 因此, 有

$$v + v_0 = \sum_{i=0}^{m} \lambda_i v_i, \quad \sum_{i=0}^{m} \lambda_i = 1.$$

这蕴涵着关系

$$v = \sum_{i=1}^{m} \lambda_i (v_i - v_0) \in \operatorname{span}\{v_i - v_0 \mid i = 1, \cdots, m\},$$

这就证明了相反的包含关系. □

下面命题的证明是相当直截了当的.

命题 1.65 元素 $v_0, \cdots, v_m$ 在 $\mathbb{R}^n$ 中是仿射无关的当且仅当它的仿射包 $\Omega := \operatorname{aff}\{v_0, \cdots, v_m\}$ 是 m-维的.

证明　假设 $v_0, \cdots, v_m$ 是仿射无关的, 则由引理 1.64 得子空间 $L := \operatorname{span}\{v_i - v_0 \mid i = 1, \cdots, m\}$ 平行于 Ω. 根据命题 1.63, $v_1 - v_0, \cdots, v_m - v_0$ 线性无关, 从而子空间 L 是 m-维的, 因此 Ω 也是 m-维的. 相反的结论的证明也是容易的. □

仿射无关系统可引出单纯型的构造.

定义 1.66　设 $v_0, \cdots, v_m$ 在 $\mathbb{R}^n$ 中是仿射无关的, 则集合

$$\Delta_m := \operatorname{co}\{v_i \mid i = 0, \cdots, m\}$$

称为 $\mathbb{R}^n$ 中以 $v_i\,(i = 0, \cdots, m)$ 为顶点的 m-单纯型.

单纯型顶点的重要作用由下面的命题揭示.

命题 1.67　考虑顶点为 $v_i(i = 0, \cdots, m)$ 的 m-单纯型 Δ_m, 则对任意的 $v \in \Delta_m$, 存在唯一的元素 $(\lambda_0, \cdots, \lambda_m) \in \mathbb{R}^{m+1}_+$, 使得

$$v = \sum_{i=0}^{m} \lambda_i v_i, \quad \sum_{i=0}^{m} \lambda_i = 1.$$

证明　设 $(\lambda_0, \cdots, \lambda_m) \in \mathbb{R}^{m+1}_+$, $(\mu_0, \cdots, \mu_m) \in \mathbb{R}^{m+1}_+$ 满足

$$v = \sum_{i=0}^{m} \lambda_i v_i = \sum_{i=0}^{m} \mu_i v_i, \quad \sum_{i=0}^{m} \lambda_i = \sum_{i=0}^{m} \mu_i = 1,$$

则

$$\sum_{i=0}^{m} (\lambda_i - \mu_i) v_i = 0, \quad \sum_{i=0}^{m} (\lambda_i - \mu_i) = 0.$$

由于 $v_0, \cdots, v_m$ 是仿射无关的, 得 $\lambda_i = \mu_i$, $i = 0, \cdots, m$. □

现在我们已准备好来定义凸集的相对内部这个重要概念.

定义 1.68　设 Ω 是凸集, 称元素 $v \in \Omega$ 属于 Ω 的相对内部 $\operatorname{ri}\Omega$, 如果存在 $\varepsilon > 0$, 使得 $\mathbb{B}(v; \varepsilon) \wedge \operatorname{aff}\Omega \subset \Omega$.

我们从下面的引理开始相对内部的研究.

引理 1.69　任何线性映射 $A : \mathbb{R}^n \to \mathbb{R}^p$ 是连续的.

证明　设 $\{e_1, \cdots, e_n\}$ 是 $\mathbb{R}^n$ 的标准正交基, 设 $v_i := A(e_i), i = 1, \cdots, m$. 对任意的 $x \in \mathbb{R}^n$ 且 $x = (x_1, \cdots, x_n)$, 有

$$A(x) = A\left(\sum_{i=1}^{n} x_i e_i\right) = \sum_{i=1}^{n} x_i A(e_i) = \sum_{i=1}^{n} x_i v_i.$$

由三角不等式和 Cauchy-Schwarz 不等式得

$$\|A(x)\| \leqslant \sum_{i=1}^{n} |x_i|\|v_i\| \leqslant \sqrt{\sum_{i=1}^{n} |x_i|^2}\sqrt{\sum_{i=1}^{n} \|v_i\|^2} = M\|x\|,$$

其中 $M := \sqrt{\sum_{i=1}^{n} \|v_i\|^2}$. 进一步可得

$$\|A(x) - A(y)\| = \|A(x-y)\| \leqslant M\|x-y\|, \quad \forall\, x, y \in \mathbb{R}^n,$$

这就蕴涵着映射 A 的连续性. □

下面的命题在本书中起着至关重要的作用.

命题 1.70 设 Δ_m 是 $\mathbb{R}^n$ 中的 m-单纯型, 其中 $m \geqslant 1$, 则 $\operatorname{ri}\Delta_m \neq \varnothing$.

证明 考虑单纯型 Δ_m 的顶点 $v_0, \cdots, v_m$, 记

$$v := \frac{1}{m+1}\sum_{i=0}^{m} v_i.$$

通过证明 $v \in \operatorname{ri}\Delta_m$ 来证明命题成立. 定义

$$L := \operatorname{span}\{v_i - v_0 \mid i = 1, \cdots, m\},$$

并注意到 L 是平行于 $\operatorname{aff}\Delta_m = \operatorname{aff}\{v_0, \cdots, v_m\}$ 的 $\mathbb{R}^n$ 的 m-维子空间. 易见对任意的 $x \in L$, 存在唯一的 $(\lambda_0, \cdots, \lambda_m) \in \mathbb{R}^{m+1}$ 满足

$$x = \sum_{i=0}^{m} \lambda_i v_i, \quad \sum_{i=0}^{m} \lambda_i = 0.$$

考虑映射 $A : L \to \mathbb{R}^{m+1}$, 它映 x 到如上相应的系数 $(\lambda_0, \cdots, \lambda_m) \in \mathbb{R}^{m+1}$, 则 A 是线性的, 因此根据引理 1.69 它是连续的. 由于 $A(0) = 0$, 可选取 $\delta > 0$ 使得

$$\|A(u)\| < \frac{1}{m+1} \text{ 对任意的 } \|u\| \leqslant \delta\text{成立}.$$

现证 $(v + \delta\mathbb{B}) \cap \operatorname{aff}\Delta_m \subset \Delta_m$, 从而 $v \in \operatorname{ri}\Delta_m$. 为此, 任取 $x \in (v + \delta\mathbb{B}) \cap \operatorname{aff}\Delta_m$, 则对某 $u \in \delta\mathbb{B}$ 有 $x = v + u$. 由于 v, $x \in \operatorname{aff}\Delta_m$, $u = x - v$, 有 $u \in L$. 记 $A(u) := (\alpha_0, \cdots, \alpha_m)$, 则 $u = \sum_{i=0}^{m} \alpha_i v_i$, 其中 $\sum_{i=0}^{m} \alpha_i = 0$ 且有估计

$$|\alpha_i| \leqslant \|A(u)\| < \frac{1}{m+1}, \quad \forall\, i = 0, \cdots, m,$$

则相应地蕴涵着

$$v+u=\sum_{i=0}^{m}\left(\frac{1}{m+1}+\alpha_i\right)v_i=\sum_{i=0}^{m}\mu_i v_i,$$

这里, $\mu_i:=\dfrac{1}{m+1}+\alpha_i\geqslant 0,\ i=1,\cdots,m$. 由于 $\sum_{i=0}^{m}\mu_i=1$, 这就保证 $x\in\Delta_m$. 因此 $(v+\delta\mathbb{B})\cap\operatorname{aff}\Delta_m\subset\Delta_m$, 从而 $v\in\operatorname{ri}\Delta_m$. □

引理 1.71　设 Ω 是 $\mathbb{R}^n$ 中维数 $m\geqslant 1$ 的非空凸集, 则在 Ω 中存在 $m+1$ 个仿射无关的元素 $v_0,\cdots,v_m$.

证明　设 $\Delta_k:=\operatorname{co}\{v_0,\cdots,v_k\}$ 是包含在 Ω 内的最大维数的 k-单纯型, 则 $v_0,\cdots,v_k$ 是仿射无关的. 现为证 $k=m$, 构造 $K:=\operatorname{aff}\{v_0,\cdots,v_k\}$, 并注意到由于 $\{v_0,\cdots,v_k\}\subset\Omega$, 故有 $K\subset\operatorname{aff}\Omega$. 接下来可证明 $\Omega\subset K$, 所以相反的包含关系也成立. 为此, 用反证法证明. 假设存在 $\omega\in\Omega$ 使得 $\omega\notin K$, 则直接利用仿射无关的定义可证, 作为 Ω 的子集, $\{v_0,\cdots,v_k,\omega\}$ 是仿射无关的, 这是一个矛盾. 因此 $K=\operatorname{aff}\Omega$, 从而得 $k=\dim K=\dim\operatorname{aff}\Omega=\dim\Omega=m$. □

下面的定理是凸有限维几何中最基本的结果之一.

定理 1.72　设 Ω 是 $\mathbb{R}^n$ 的非空凸集, 则下面的结论成立:

(i) 总有 $\operatorname{ri}\Omega\neq\varnothing$.

(ii) 对任意的 $a\in\operatorname{ri}\Omega$, $b\in\overline{\Omega}$, 有 $[a,b)\subset\operatorname{ri}\Omega$.

证明　(i) 设 m 是 Ω 的维数. 首先注意到 $m=0$ 的情形是平凡的, 由于此时 Ω 是单点集而且 $\operatorname{ri}\Omega=\Omega$. 假设 $m\geqslant 1$, 并由引理 1.71 在 Ω 中找到 $m+1$ 个仿射无关的元素 $v_0,\cdots,v_m$. 进一步考虑 m-单纯型

$$\Delta_m:=\operatorname{co}\{v_0,\cdots,v_m\}.$$

可证明 $\operatorname{aff}\Delta_m=\operatorname{aff}\Omega$. 为完成证明, 取 $v\in\operatorname{ri}\Delta_m$, 根据命题 1.70, 存在对任意小的 $\varepsilon>0$, 有

$$\mathbb{B}(v;\varepsilon)\cap\operatorname{aff}\Omega=\mathbb{B}(v;\varepsilon)\cap\operatorname{aff}\Delta_m\subset\Delta_m\subset\Omega.$$

根据相对内部的定义这就证明了 $v\in\operatorname{ri}\Omega$.

(ii) 设 L 是平行于 $\operatorname{aff}\Omega$ 的 $\mathbb{R}^n$ 的子空间, 令 $m:=\dim L$, 则存在双射线性映射 $A:L\to\mathbb{R}^m$ 使得 A 和 A^{-1} 都是连续的. 取定 $x_0\in\operatorname{aff}\Omega$, 并定义映射 $f:\operatorname{aff}\Omega\to\mathbb{R}^m$ 为 $f(x):=A(x-x_0)$. 易验证 f 是双射仿射映射, 而且 f 和 f^{-1} 都是连续的. 我们还看到 $a\in\operatorname{ri}\Omega$ 当且仅当 $f(a)\in\operatorname{int}f(\Omega)$, $b\in\overline{\Omega}$ 当且仅当 $f(b)\in\overline{f(\Omega)}$, 则根据引理 1.30, 有 $[f(a),f(b))\subset\operatorname{int}f(\Omega)$. 这就证明了 $[a,b)\subset\operatorname{ri}\Omega$. □

最后介绍相对内部的性质.

命题 1.73 设 Ω 是 $\mathbb{R}^n$ 的非空凸子集, 对凸集 $\operatorname{ri}\Omega$ 和 $\overline{\Omega}$, 有

(i) $\overline{\operatorname{ri}\Omega} = \overline{\Omega}$;

(ii) $\operatorname{ri}\Omega = \operatorname{ri}\overline{\Omega}$.

证明 注意到 $\operatorname{ri}\Omega$ 的凸性由定理 1.72(ii) 可得, 而 $\overline{\Omega}$ 的凸性在命题 1.29 中已证明. 为证此命题的结论 (i) 成立, 注意到包含关系 $\overline{\operatorname{ri}\Omega} \subset \overline{\Omega}$ 是显然的. 取定 $b \in \overline{\Omega}$, 选取 $a \in \operatorname{ri}\Omega$, 构造序列

$$x_k := \frac{1}{k}a + \left(1 - \frac{1}{k}\right)b, \quad k \in \mathbb{N},$$

当 $k \to \infty$ 时收敛于 b. 由于根据定理 1.72(ii), $x_k \in \operatorname{ri}\Omega$, 所以有 $b \in \overline{\operatorname{ri}\Omega}$. 从而 $\overline{\Omega} \subset \overline{\operatorname{ri}\Omega}$, 这就证明了 (i) 成立. (ii) 的证明类似于命题 1.31(ii) 的证明. □

1.5 距离函数

本节主要研究一类凸集的距离函数, 它是凸分析及其推广的最有趣和重要的课题. 这类函数本质上是非可微的, 因此自然和频繁地出现在分析及其应用中.

给定集合 $\Omega \subset \mathbb{R}^n$, 关联于 Ω 的距离函数 (图 1.6) 定义为

$$d(x;\Omega) := \inf\{\|x - \omega\| \mid \omega \in \Omega\}. \tag{1.5}$$

进一步回顾映射 $f : \mathbb{R}^n \to \mathbb{R}^p$ 在某集合 $C \subset \mathbb{R}^n$ 上是具有常数 $\ell \geqslant 0$ 的 Lipschitz 连续的, 如果有估计

$$\|f(x) - f(y)\| \leqslant \ell\|x - y\|, \quad \forall x, y \in C. \tag{1.6}$$

注意到估计 (1.6) 中 f 的 Lipschitz 连续性说明它的连续性具有线性率.

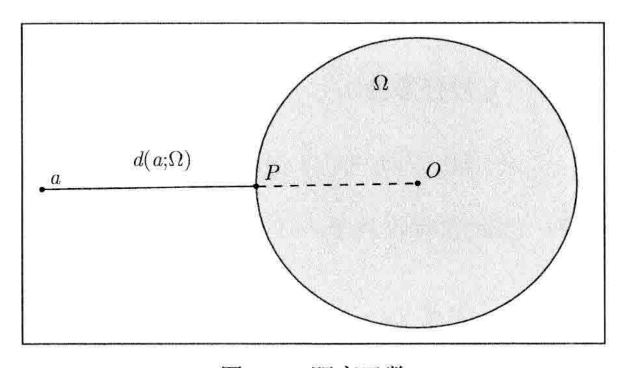


图 1.6 距离函数

命题 1.74　设 Ω 是 $\mathbb{R}^n$ 的非空子集, 则下面的结论成立:

(i) $d(x;\Omega)=0$ 当且仅当 $x\in\overline{\Omega}$.

(ii) 函数 $d(x;\Omega)$ 在 $\mathbb{R}^n$ 上是具常数 $\ell=1$ 的 Lipschitz 连续的.

证明　(i) 假设 $d(x;\Omega)=0$, 对任意的 $k\in\mathbb{N}$, 找到 $\omega_k\in\Omega$ 使得

$$0=d(x;\Omega)\leqslant\|x-\omega_k\|<d(x;\Omega)+\frac{1}{k}=\frac{1}{k},$$

这保证序列 $\{\omega_k\}$ 收敛于 x, 因此 $x\in\overline{\Omega}$.

反之, 设 $x\in\overline{\Omega}$, 并找到收敛于 x 的序列 $\{\omega_k\}\subset\Omega$, 则

$$0\leqslant d(x;\Omega)\leqslant\|x-\omega_k\|,\quad\forall k\in\mathbb{N},$$

由于 $\|x-\omega_k\|\to 0\,(k\to\infty)$, 这蕴涵着 $d(x;\Omega)=0$.

(ii) 对任意的 $\omega\in\Omega$, 有估计

$$d(x;\Omega)\leqslant\|x-\omega\|\leqslant\|x-y\|+\|y-\omega\|,$$

它相应地蕴涵着

$$d(x;\Omega)\leqslant\|x-y\|+\inf\{\|y-\omega\|\mid\omega\in\Omega\}=\|x-y\|+d(y;\Omega).$$

同理, 有 $d(x;\Omega)\leqslant\|y-x\|+d(x;\Omega)$, 因此 $|d(x;\Omega)-d(x;\Omega)|\leqslant\|x-y\|$, 根据式 (1.6), 这就证明了 $d(x;\Omega)$ 在 $\mathbb{R}^n$ 上具有常数 $\ell=1$ 的 Lipschitz 连续性. □

对任意的 $x\in\mathbb{R}^n$, x 到 Ω 的 Euclid 投影定义为

$$\Pi(x;\Omega):=\{\omega\in\Omega\mid\|x-\omega\|=d(x;\Omega)\}.\tag{1.7}$$

命题 1.75　设 Ω 是 $\mathbb{R}^n$ 的非空闭子集, 则对任意的 $x\in\mathbb{R}^n$, Euclid 投影 $\Pi(x;\Omega)$ 是非空的.

证明　根据定义 (1.7), 对任意的 $k\in\mathbb{N}$, 存在 $\omega_k\in\Omega$ 使得

$$d(x;\Omega)\leqslant\|x-\omega_k\|<d(x;\Omega)+\frac{1}{k}.$$

显然 $\{\omega_k\}$ 是有界序列. 因此它有收敛于 ω 的子序列 $\{\omega_{k_\ell}\}$. 由于 Ω 是闭的, 所以 $\omega\in\Omega$. 在不等式

$$d(x;\Omega)\leqslant\|x-\omega_{k_\ell}\|<d(x;\Omega)+\frac{1}{k_\ell}$$

中令 $\ell\to\infty$, 有 $d(x;\Omega)=\|x-\omega\|$, 这就保证 $\omega\in\Pi(x;\Omega)$. □

一个有趣的凸性结果是下面的唯一投影性质.

推论 1.76 如果 Ω 是 $\mathbb{R}^n$ 的非空闭凸子集, 则对任意的 $x \in \mathbb{R}^n$, Euclid 投影 $\Pi(x;\Omega)$ 是单点集.

证明 投影 $\Pi(x;\Omega)$ 的非空性由命题 1.75 直接可得. 为证唯一性, 假设 $\omega_1, \omega_2 \in \Pi(x;\Omega)$ 且 $\omega_1 \neq \omega_2$, 则

$$\|x-\omega_1\| = \|x-\omega_2\| = d(x;\Omega).$$

根据经典的平行四边形等式, 有

$$2\|x-\omega_1\|^2 = \|x-\omega_1\|^2 + \|x-\omega_2\|^2 = 2\left\|x-\frac{\omega_1+\omega_2}{2}\right\|^2 + \frac{\|\omega_1+\omega_2\|^2}{2}.$$

这直接蕴涵着

$$\left\|x-\frac{\omega_1+\omega_2}{2}\right\|^2 = \|x-\omega_1\|^2 - \frac{\|\omega_1-\omega_2\|^2}{4} < \|x-\omega_1\|^2 = [d(x;\Omega)]^2,$$

由于 $\dfrac{\omega_1+\omega_2}{2} \in \Omega$, 这是一个矛盾. □

现在证明非空闭集的凸性与它的距离函数的凸性是等价的. 易证任意集合 Ω 的凸性蕴涵着它的距离函数的凸性.

命题 1.77 设 Ω 是 $\mathbb{R}^n$ 的非空闭子集, 则函数 $d(\cdot;\Omega)$ 是凸的当且仅当集合 Ω 是凸的.

证明 假设 Ω 是凸的, 取 $x_1, x_2 \in \mathbb{R}^n$, $\omega_i := \Pi(x_i;\Omega)$, 有

$$\|x_i-\omega_i\| = d(x_i;\Omega), \quad i=1,2.$$

Ω 的凸性保证对任意的 $\lambda \in (0,1)$, $\lambda\omega_1+(1-\lambda)\omega_2 \in \Omega$. 因此有

$$\begin{aligned} d(\lambda x_1+(1-\lambda)x_2;\Omega) &\leqslant \|\lambda x_1+(1-\lambda)x_2-[\lambda\omega_1+(1-\lambda)\omega_2]\| \\ &\leqslant \lambda\|x_1-\omega_1\|+(1-\lambda)\|x_2-\omega_2\| \\ &= \lambda d(x_1;\Omega)+(1-\lambda)d(x_2;\Omega), \end{aligned}$$

这蕴涵着距离函数 $d(\cdot;\Omega)$ 的凸性.

为证反向关系成立, 假设 $d(\cdot;\Omega)$ 是凸的, 任取 $\omega_i \in \Omega$, $i=1,2$, $\lambda \in (0,1)$, 则有

$$d(\lambda\omega_1+(1-\lambda)\omega_2;\Omega) \leqslant \lambda d(\omega_1;\Omega)+(1-\lambda)d(\omega_2;\Omega) = 0.$$

由于 Ω 是闭的, 所以有 $\lambda\omega_1+(1-\lambda)\omega_2 \in \Omega$, 这样就证明了 Ω 的凸性. □

接下来刻画到 $\mathbb{R}^n$ 中凸集的 Euclid 投影. 在下面的命题和之后的讨论中, 如果 Ω 是非空闭凸集, 我们也用投影 $\Pi(x;\Omega)$ 来表示其唯一的元素.

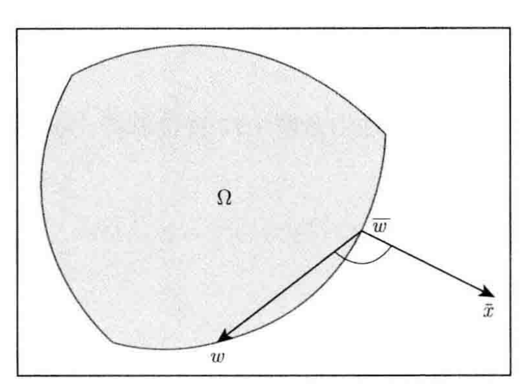


图 1.7　Euclid 投影

命题 1.78　设 Ω 是 $\mathbb{R}^n$ 的非空凸子集, $\bar{\omega} \in \Omega$, 则 $\bar{\omega} \in \Pi(\bar{x};\Omega)$ 当且仅当

$$\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega} \rangle \leqslant 0, \quad \forall \omega \in \Omega. \tag{1.8}$$

证明　取 $\bar{\omega} \in \Pi(\bar{x};\Omega)$, 对任意的 $\omega \in \Omega, \lambda \in (0,1)$, 有 $\bar{\omega} + \lambda(\omega - \bar{\omega}) \in \Omega$. 因此

$$\begin{aligned}\|\bar{x} - \bar{\omega}\|^2 &= [d(\bar{x};\Omega)]^2] \leqslant \|\bar{x} - [\bar{\omega} + \lambda(\omega - \bar{\omega})]\|^2 \\ &= \|\bar{x} - \bar{\omega}\|^2 - 2\lambda\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega} \rangle + \lambda^2\|\omega - \bar{\omega}\|^2,\end{aligned}$$

这已经蕴涵着

$$2\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega} \rangle \leqslant \lambda\|\omega - \bar{\omega}\|^2.$$

令 $\lambda \to 0^+$, 得式 (1.8) 成立.

为证反向结论成立, 假设式 (1.8) 成立, 则对任意的 $\omega \in \Omega$, 有

$$\begin{aligned}\|\bar{x} - \omega\|^2 &= \|\bar{x} - \bar{\omega} + \bar{\omega} - \omega\|^2 \\ &= \|\bar{x} - \bar{\omega}\|^2 + \|\bar{\omega} - \omega\|^2 + 2\langle \bar{x} - \bar{\omega}, \bar{\omega} - \omega \rangle \\ &= \|\bar{x} - \bar{\omega}\|^2 + \|\bar{\omega} - \omega\|^2 - 2\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega} \rangle \geqslant \|\bar{x} - \bar{\omega}\|^2.\end{aligned}$$

因此 $\|\bar{x} - \bar{\omega}\| \leqslant \|\bar{x} - \omega\|, \forall \omega \in \Omega$, 这蕴涵着 $\bar{\omega} \in \Pi(\bar{x};\Omega)$, 从而完成了证明. □

从上面的结果我们知道, 对 $\mathbb{R}^n$ 中的任意非空闭凸集 Ω, Euclid 投影 $\Pi(x;\Omega)$ 是单点集. 现在, 证明投影映射事实上是非扩张的, 即它满足下面的 Lipschitz 性质.

命题 1.79　设 Ω 是 $\mathbb{R}^n$ 的非空闭凸子集, 则对任意元素 $x_1, x_2 \in \mathbb{R}^n$, 有估计

$$\|\Pi(x_1;\Omega) - \Pi(x_2;\Omega)\|^2 \leqslant \langle \Pi(x_1;\Omega) - \Pi(x_2;\Omega), x_1 - x_2 \rangle.$$

特别地, 它蕴涵着投影具常数 $\ell = 1$ 的 Lipschitz 连续性:

$$\|\Pi(x_1;\Omega) - \Pi(x_2;\Omega)\| \leqslant \|x_1 - x_2\|, \quad \forall x_1, x_2 \in \mathbb{R}^n.$$

证明　由命题 1.78 可得

$$\langle \Pi(x_2;\Omega)-\Pi(x_1;\Omega), x_1-\Pi(x_1;\Omega)\rangle \leqslant 0, \quad \forall x_1,x_2\in\mathbb{R}^n.$$

在上面的不等式中对换 x_1,x_2, 并把它们相加得

$$\langle \Pi(x_1;\Omega)-\Pi(x_2;\Omega), x_2-x_1+\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\rangle \leqslant 0.$$

这蕴涵着命题中第一个估计式成立. 最后, Euclid 投影的非扩张性质由下式直接可得

$$\begin{aligned}\|\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\|^2 &\leqslant \langle \Pi(x_1;\Omega)-\Pi(x_2;\Omega), x_1-x_2\rangle\\ &\leqslant \|\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\|\cdot\|x_1-x_2\|(\forall x_1,x_2\in\mathbb{R}^n).\end{aligned}$$

这就完成了命题的证明. □

1.6 练　　习

练习 1.1　设 Ω_1 和 Ω_2 是 $\mathbb{R}^n$ 的非空闭凸子集, 满足 Ω_1 是有界的或 Ω_2 是有界的. 证明 $\Omega_1-\Omega_2$ 是非空闭凸集. 举一个例子: Ω_1,Ω_2 是两个非空闭凸集, 但 $\Omega_1-\Omega_2$ 不是闭的.

练习 1.2　设 Ω 是 $\mathbb{R}^n$ 的子集, 称 Ω 是锥, 如果 $\lambda x\in\Omega, \forall\lambda\geqslant 0, \forall x\in\Omega$. 证明以下结论是等价的:

(i) Ω 是凸锥,

(ii) $x+y\in\Omega, \forall x,y\in\Omega$, 而且 $\lambda x\in\Omega, \forall x\in\Omega, \forall\lambda\geqslant 0$.

练习 1.3　(i) 设 Ω 是含 0 的非空凸集, 且 $0\leqslant\lambda_1\leqslant\lambda_2$, 证明 $\lambda_1\Omega\subset\lambda_2\Omega$.

(ii) 设 Ω 是非空凸集, 且 $\alpha,\beta\geqslant 0$, 证明 $\alpha\Omega+\beta\Omega\subset(\alpha+\beta)\Omega$.

练习 1.4　(i) 设 $\Omega_i(i=1,\cdots,m)$ 是 $\mathbb{R}^n$ 的非空凸集. 证明 $x\in \mathrm{co}\bigcup_{i=1}^m\Omega_i$ 当且仅当存在元素 $\omega_i\in\Omega_i$, $\lambda_i\geqslant 0$, $i=1,\cdots,m$ 且 $\sum\limits_{i=1}^m\lambda_i=1$ 使得 $x=\sum\limits_{i=1}^m\lambda_i\omega_i$.

(ii) 设 $\Omega_i(i=1,\cdots,m)$ 是 $\mathbb{R}^n$ 的非空凸锥. 证明

$$\sum_{i=1}^m\Omega_i=\mathrm{co}\left\{\bigcup_{i=1}^m\Omega_i\right\}.$$

练习 1.5　设 Ω 是 $\mathbb{R}^n$ 的非空凸锥. 证明 Ω 是 $\mathbb{R}^n$ 的线性子空间当且仅当 $\Omega=-\Omega$.

练习 1.6　证明下面的函数在 $\mathbb{R}^n$ 上是凸的:

(i) $f(x)=\alpha\|x\|$, 其中 $\alpha\geqslant 0$.

(ii) $f(x) = \|x - a\|^2$, 其中 $a \in \mathbb{R}^n$.

(iii) $f(x) = \|Ax - b\|$, 其中 A 是 $p \times n$ 矩阵, $b \in \mathbb{R}^p$.

(iv) $f(x) = \|x\|^q$, 其中 $q \geqslant 1$.

练习 1.7　证明下面的函数在其定义域上是凸的:

(i) $f(x) = \mathrm{e}^{ax}$, $x \in \mathbb{R}$, 这里 a 是常数.

(ii) $f(x) = x^q$, $x \in [0, \infty)$, 其中 $q \geqslant 1$ 是常数.

(iii) $f(x) = -\ln(x)$, $x \in (0, \infty)$.

(iv) $f(x) = x\ln(x)$, $x \in (0, \infty)$.

练习 1.8　(i) 举一个例子: 函数 $f : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$, 它相对于每个变量是凸的, 但作为两个变量的函数却不是凸的.

(ii) 设 $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}\,(i = 1, 2)$ 是凸函数. 你能确定最小值函数

$$\min\{f_1, f_2\}(x) := \min\{f_1(x), f_2(x)\}$$

是凸的吗?

练习 1.9　举例说明两个实值凸函数的乘积不一定是凸的.

练习 1.10　证明集合 $\{(x, t) \in \mathbb{R}^n \times \mathbb{R} \mid t \geqslant \|x\|\}$ 是闭凸的.

练习 1.11　关联于集合 Ω 的指示函数定义为

$$\delta(x; \Omega) := \begin{cases} 0, & \text{如果 } x \in \Omega, \\ \infty, & \text{其他}. \end{cases}$$

(i) 计算当 $\Omega = \varnothing$, $\Omega = \mathbb{R}^n$ 和 $\Omega = [-1, 1]$ 时的 $\delta(\cdot; \Omega)$.

(ii) 证明集合 Ω 是凸集当且仅当它的指示函数 $\delta(\cdot; \Omega)$ 是凸的.

练习 1.12　证明如果 $f : \mathbb{R} \to [0, \infty)$ 是凸函数, 那么它的 q 次幂 $f^q(x) := [f(x)]^q\,(q \geqslant 1)$ 也是凸的.

练习 1.13　设 $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, 2$ 是凸函数. 定义

$$\Omega_1 := \{(x, \lambda_1, \lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid \lambda_1 \geqslant f_1(x)\},$$

$$\Omega_2 := \{(x, \lambda_1, \lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid \lambda_2 \geqslant f_2(x)\}.$$

(i) 证明集合 Ω_1, Ω_2 是凸的.

(ii) 定义集值映射 $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^2$ 为 $F(x) := [f_1(x), \infty) \times [f_2(x), \infty)$, 证明 F 的图是 $\Omega_1 \cap \Omega_2$.

练习 1.14　称 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是正齐次的, 如果 $f(\alpha x) = \alpha f(x)$, $\forall\, \alpha > 0$, 称 f 是次可加的, 如果 $f(x + y) \leqslant f(x) + f(y)$, $\forall\, x, y \in \mathbb{R}^n$. 证明正齐次函数是凸的当且仅当它是次可加的.

练习 1.15　给定非空集合 $\Omega \subset \mathbb{R}^n$, 定义

$$K_\Omega := \{\lambda x \mid \lambda \geqslant 0,\ x \in \Omega\} = \bigcup_{\lambda \geqslant 0} \lambda\Omega.$$

(i) 证明 K_Ω 是锥.

(ii) 证明 K_Ω 是包含 Ω 的最小的锥.

(iii) 证明如果 Ω 是凸的, 那么锥 K_Ω 也是凸的.

练习 1.16　设 $\varphi : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$ 是凸函数, 设 K 是 $\mathbb{R}^p$ 的非空凸子集. 假设对任意的 $x \in \mathbb{R}^n$, 函数 $\varphi(x, \cdot)$ 在 K 上是下有界的. 证明在 $\mathbb{R}^n$ 上定义的函数 $f(x) := \inf\{\varphi(x, y) \mid y \in K\}$ 是凸的.

练习 1.17　设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸函数.

(i) 证明对任意的 $\alpha \in \mathbb{R}$, 水平集 $\{x \in \mathbb{R}^n \mid f(x) \leqslant \alpha\}$ 是凸的.

(ii) 设 $\Omega \in \overline{\mathbb{R}}$ 是凸集, 逆像 $f^{-1}(\Omega)$ 一定是 $\mathbb{R}^n$ 的凸子集吗?

练习 1.18　设 C 是 $\mathbb{R}^{n+1}$ 的凸子集.

(i) 对 $x \in \mathbb{R}^n$, 定义 $F(x) := \{\lambda \in \mathbb{R} \mid (x, \lambda) \in C\}$. 证明 F 是具有凸图的集值映射, 并给出它的图的显式公式.

(ii) 对 $x \in \mathbb{R}^n$, 定义函数

$$f_C(x) := \inf\{\lambda \in \mathbb{R} \mid (x, \lambda) \in C\}. \tag{1.9}$$

当 C 是 $\mathbb{R}^2$ 的闭单位球时找到 f_C 的一个显式公式.

(iii) 对于 (ii) 中定义的函数 f_C, 证明如果对任意的 $x \in \mathbb{R}^n, f_C(x) > -\infty$, 那么 f_C 是凸函数.

练习 1.19　设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是下有界的. 定义它的凸化为

$$(\operatorname{co} f)(x) := \inf\left\{\sum_{i=1}^m \lambda_i f(x_i) \,\middle|\, \lambda_i \geqslant 0, \sum_{i=1}^m \lambda_i = 1, \sum_{i=1}^m \lambda_i x_i = x,\ m \in \mathbb{N}\right\}. \tag{1.10}$$

证明式 (1.10) 是凸的且 $\operatorname{co} f = f_C$, 其中 $C := \operatorname{co}(\operatorname{epi} f), f_C$ 在式 (1.9) 中定义.

练习 1.20　称 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是拟凸的, 如果

$$f(\lambda x + (1-\lambda)y) \leqslant \max\{f(x), f(y)\}, \quad \forall, x, y \in \mathbb{R}^n, \quad \lambda \in (0,1).$$

(i) 证明函数 f 是拟凸的当且仅当对任意的 $\alpha \in \mathbb{R}$, 水平集 $\{x \in \mathbb{R}^n \mid f(x) \leqslant \alpha\}$ 是凸集.

(ii) 证明任何凸函数 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是拟凸的. 举一个例子说明反之不成立.

练习 1.21　设 a 是 $\mathbb{R}^n$ 中的非零元, $b \in \mathbb{R}$. 证明

$$\Omega := \{x \in \mathbb{R}^n \mid \langle a, x\rangle = b\}$$

是仿射集且 $\dim\Omega = n-1$.

练习 1.22　设集合 $\{v_1, \cdots, v_m\}$ 由仿射无关元素组成, 且 $v \notin \operatorname{aff}\{v_1, \cdots, v_m\}$, 证明 $v_1, \cdots, v_m, v$ 是仿射无关的.

练习 1.23　假设 Ω 是 $\mathbb{R}^n$ 的凸子集当 $m \geqslant 1$, 有 $\dim\Omega = m$. 设集合 $\{v_1, \cdots, v_m\} \subset \Omega$ 是由仿射无关的元素组成的, 并令

$$\Delta_m := \operatorname{co}\{v_1, \cdots, v_m\}.$$

证明 $\operatorname{aff}\Omega = \operatorname{aff}\Delta_m = \operatorname{aff}\{v_1, \cdots, v_m\}$.

练习 1.24　设 Ω 是 $\mathbb{R}^n$ 的非空凸子集. 证明:

(i) $\operatorname{aff}\Omega = \operatorname{aff}\overline{\Omega}$.

(ii) 对任意的 $\bar{x} \in \operatorname{ri}\Omega, x \in \overline{\Omega}$, 存在 $t > 0$ 使得 $\bar{x} + t(\bar{x} - x) \in \Omega$.

(iii) 命题 1.73(ii).

练习 1.25　设 $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ 是凸集且 $\operatorname{ri}\Omega_1 \cap \operatorname{ri}\Omega_2 \neq \varnothing$. 证明:

(i) $\overline{\Omega_1 \cap \Omega_2} = \overline{\Omega}_1 \cap \overline{\Omega}_2$.

(ii) $\operatorname{ri}(\Omega_1 \cap \Omega_2) = \operatorname{ri}\Omega_1 \cap \operatorname{ri}\Omega_2$.

练习 1.26　设 $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ 是凸集且 $\overline{\Omega}_1 = \overline{\Omega}_2$. 证明 $\operatorname{ri}\Omega_1 = \operatorname{ri}\Omega_2$.

练习 1.27　(i) 设 $B : \mathbb{R}^n \to \mathbb{R}^p$ 是仿射映射, 设 Ω 是 $\mathbb{R}^n$ 的凸子集. 证明等式

$$B(\operatorname{ri}\Omega) = \operatorname{ri}B(\Omega).$$

(ii) 设 Ω_1 和 Ω_2 是 $\mathbb{R}^n$ 的凸子集, 证明 $\operatorname{ri}(\Omega_1 - \Omega_2) = \operatorname{ri}\Omega_1 - \operatorname{ri}\Omega_2$.

练习 1.28　设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸函数. 证明:

(i) $\operatorname{aff}(\operatorname{epi}f) = \operatorname{aff}(\operatorname{dom}f) \times \mathbb{R}$.

(ii) $\operatorname{ri}(\operatorname{epi}f) = \{(x, \lambda) \mid x \in \operatorname{ri}(\operatorname{dom}f), \lambda > f(x)\}$.

练习 1.29　找出下面距离函数 $d(x;\Omega)$ 和 Euclid 投影 $\Pi(x;\Omega)$ 的显式公式:

(i) Ω 是 $\mathbb{R}^n$ 的闭单位球.

(ii) $\Omega := \{(x_1, x_2) \in \mathbb{R}^2 \mid |x_1| + |x_2| \leqslant 1\}$.

(iii) $\Omega := [-1, 1] \times [-1, 1]$.

练习 1.30　找到下面的投影 $\Pi(x;\Omega)$ 的公式:

(i) $\Omega := \{x \in \mathbb{R}^n \mid \langle a, x\rangle = b\}$, 其中 $0 \neq a \in \mathbb{R}^n$, $b \in \mathbb{R}$.

(ii) $\Omega := \{x \in \mathbb{R}^n \mid Ax = b\}$, 其中 A 是 $m \times n$ 矩阵且 $\operatorname{rank}A = m$, $b \in \mathbb{R}^m$.

(iii) Ω 是非负象限 $\Omega := \mathbb{R}^n_+$.

练习 1.31　对 $a_i \leqslant b_i, i = 1, \cdots, n$, 定义

$$\Omega := \{x = (x_1, \cdots, x_n) \in \mathbb{R}^n \mid a_i \leqslant x_i \leqslant b_i, i = 1, \cdots, n\}.$$

证明投影 $\Pi(x;\Omega)$ 有

$$\Pi(x;\Omega) = \{u = (u_1, \cdots, u_n) \in \mathbb{R}^n \mid u_i = \max\{a_i, \min\{b, x_i\}\}, i \in \{1, \cdots, n\}\}.$$

第 2 章　次微分的运算

本章给出凸函数和凸集的广义微分理论的基本结果. 利用基于对集合的极点系统的凸分离的变分分析的几何方法, 得到集合的法向量和函数的次梯度的一般运算法则. 本章还包括关于凸函数的 Lipschitz 连续性和凸对偶性的一些相关结果. 注意到给出的运算结果的一些推广版本可在本章的练习中找到, 对它们的证明的提示将在本书书末给出.

2.1 凸　分　离

本节先介绍凸分离定理, 它在凸分析, 特别是在利用几何方法得到运算法则中起着至关重要的作用. 第一个结果是关于非空闭凸集和集合外一点的严格分离.

命题 2.1　设 Ω 是 $\mathbb{R}^n$ 的非空闭凸子集, 设 $\bar{x} \notin \Omega$, 则存在非零元素 $v \in \mathbb{R}^n$ 使得

$$\sup\{\langle v, x\rangle \mid x \in \Omega\} < \langle v, \bar{x}\rangle. \tag{2.1}$$

证明　记 $\bar{w} := \Pi(\bar{x}; \Omega)$, 令 $v := \bar{x} - \bar{w} \neq 0$. 利用命题 1.78 得

$$\langle v, x - \bar{w}\rangle = \langle \bar{x} - \bar{w}, x - \bar{w}\rangle \leqslant 0, \quad \forall x \in \Omega.$$

从而

$$\langle v, x - (\bar{x} - v)\rangle = \langle v, x - \bar{w}\rangle \leqslant 0,$$

反过来蕴涵着 $\langle v, x\rangle \leqslant \langle v, \bar{x}\rangle - \|v\|^2$, 因此证明了式 (2.1) 成立. □

在命题 2.1 的情况下, 可选取一个数 $b \in \mathbb{R}$ 使得

$$\sup\{\langle v, x\rangle \mid x \in \Omega\} < b < \langle v, \bar{x}\rangle,$$

并定义函数 $A(x) := \langle v, x\rangle$ 满足 $A(\bar{x}) > b$ 以及对所有的 $x \in \Omega, A(x) < b$, 此时则称 $\bar{x}$ 和 Ω 被超平面 $L := \{x \in \mathbb{R}^n \mid A(x) = b\}$ 严格分离.

下面的定理把命题 2.1 的结果推广到两个非空闭凸集, 其中至少有一个是有界的情形.

定理 2.2　设 Ω_1 与 Ω_2 是 $\mathbb{R}^n$ 的非空闭凸子集且 $\Omega_1 \cap \Omega_2 = \varnothing$. 如果 Ω_1 是有界的, 或 Ω_2 是有界的, 则存在非零元素 $v \in \mathbb{R}^n$ 使得

$$\sup\{\langle v, x\rangle \mid x \in \Omega_1\} < \inf\{\langle v, y\rangle \mid y \in \Omega_2\}.$$

证明　设 $\Omega := \Omega_1 - \Omega_2$, 则 Ω 是非空闭凸集且 $0 \notin \Omega$. 对 Ω 和 $\bar{x} = 0$ 应用命题 2.1, 有

$$\gamma := \sup\{\langle v, x\rangle \mid x \in \Omega\} < 0 = \langle v, \bar{x}\rangle,$$

其中 $0 \neq v \in \mathbb{R}^n$. 这样对任意的 $x \in \Omega_1, y \in \Omega_2$, 有 $\langle v, x - y\rangle \leqslant \gamma$, 因此 $\langle v, x\rangle \leqslant \gamma + \langle v, y\rangle$. 所以,

$$\sup\{\langle v, x\rangle \mid x \in \Omega_1\} \leqslant \gamma + \inf\{\langle v, y\rangle \mid y \in \Omega_2\} < \inf\{\langle v, y\rangle \mid y \in \Omega_2\},$$

这就完成了定理的证明. □

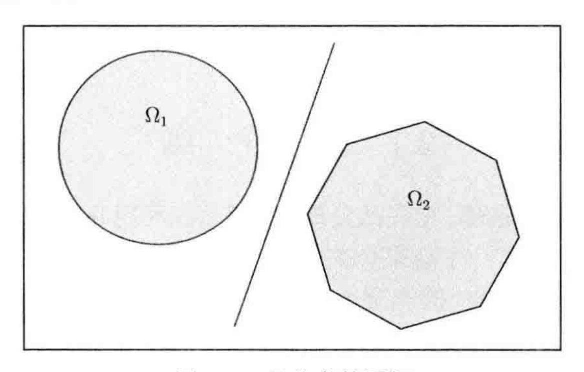


图 2.1　凸分离的图解

接下来我们证明 $\mathbb{R}^n$ 中交集为空的两个非空闭凸集合可被分离, 然而如果两个集合都是无界的, 则分离一般来说不是严格的.

推论 2.3　设 Ω_1 和 Ω_2 是 $\mathbb{R}^n$ 的非空闭凸子集满足 $\Omega_1 \cap \Omega_2 = \varnothing$, 则存在一个非零元素 $v \in \mathbb{R}^n$ 满足

$$\sup\{\langle v, x\rangle \mid x \in \Omega_1\} \leqslant \inf\{\langle v, y\rangle \mid y \in \Omega_2\}. \tag{2.2}$$

此时称这两个集是可分离的.

证明　对每个固定的 $k \in \mathbb{N}$, 定义集合 $\Theta_k := \Omega_2 \cap \mathbb{B}(0; k)$ 注意到它是闭的、有界的凸集, 如果 k 充分大, 且是非空的. 利用定理 2.2, 找到 $0 \neq u_k \in \mathbb{R}^n$ 使得

$$\langle u_k, x\rangle < \langle u_k, y\rangle, \quad \forall\, x \in \Omega_1, \quad y \in \Theta_k.$$

记 $v_k := \dfrac{u_k}{\|u_k\|}$, 不失一般性, 假设 $v_k \to v \neq 0 (k \to \infty)$. 进一步, 任取 $x \in \Omega_1, y \in \Omega_2$, 可找到 $k_0 \in \mathbb{N}$ 使得 $y \in \Theta_k$ 对任意的 $k \geqslant k_0$ 成立, 则

$$\langle v_k, x\rangle < \langle v_k, y\rangle, \quad \forall\, k \geqslant k_0.$$

令 $k \to \infty$, 有 $\langle v, x\rangle \leqslant \langle v, y\rangle$, 结论得证. □

现在将要建立凸的, 但不一定是闭的两个集合的一般分离定理. 为此, 首先证明下面的有用结果.

引理 2.4 设 Ω 是 $\mathbb{R}^n$ 的非空凸子集满足 $0 \in \overline{\Omega}\backslash\Omega$, 则存在序列 $x_k \to 0(k \to \infty)$ 且 $x_k \notin \overline{\Omega}(\forall k \in \mathbb{N})$ (图 2.2).

证明 根据定理 1.7.2(i), 有 $\text{ri}\,\Omega \neq \varnothing$. 取定 $x_0 \in \text{ri}\,\Omega$ 并证明对任意的 $t > 0$, $-tx_0 \notin \overline{\Omega}$; 参见图 2.2. 事实上, 若 $-tx_0 \in \overline{\Omega}$, 利用定理 1.72(ii) 得

$$0 = \frac{t}{1+t}x_0 + \frac{1}{1+t}(-tx_0) \in \text{ri}\,\Omega \subset \Omega$$

矛盾. 于是令 $x_k := -\dfrac{x_0}{k}$, 则对任意的 k, 有 $x_k \notin \overline{\Omega}$ 且 $x_k \to 0$. □

注意到对非凸集合引理 2.4 可能不成立. 作为一个例子, 考虑集合 $\Omega := \mathbb{R}^n\backslash\{0\}$ 满足 $0 \in \overline{\Omega}\backslash\Omega$, 然而引理 2.4 的结论不成立.

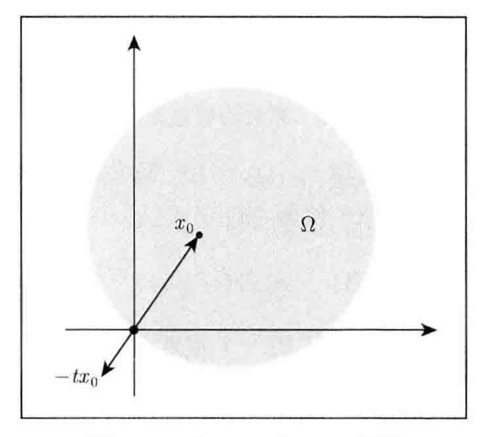


图 2.2 引理 2.4 的几何解释

定理 2.5 设 Ω_1 和 Ω_2 是 $\mathbb{R}^n$ 的非空凸子集且 $\Omega_1 \cap \Omega_2 = \varnothing$, 则对某 $v \neq 0$, 在式 (2.2) 的意义下它们是可分离的.

证明 考虑集合 $\Omega := \Omega_1 - \Omega_2$ 满足 $0 \notin \Omega$. 分别考虑以下两种情形.

情形 1 $0 \notin \overline{\Omega}$. 此时利用命题 2.1, 找到 $v \neq 0$ 使得

$$\langle v, z\rangle < 0, \quad \forall z \in \overline{\Omega}.$$

于是对任意的 $x \in \Omega_1$ 和 $y \in \Omega_2$, 有 $\langle v, x\rangle < \langle v, y\rangle$, 因此结论成立.

情形 2 $0 \in \overline{\Omega}$. 此时利用引理 2.4, 找到序列 $x_k \to 0$ 且 $x_k \notin \overline{\Omega}$, 则命题 2.1 产生一个非零元素的序列 $\{v_k\}$ 使得

$$\langle v_k, z\rangle < \langle v_k, x_k\rangle, \quad \forall z \in \overline{\Omega}.$$

不失一般性 (除以 $\|v_k\|$ 并抽取一个收敛的子序列), 可设 $v_k \to v$ 且 $\|v_k\| = \|v\| = 1$. 令 $k \to \infty$ 得

$$\langle v, z\rangle \leqslant 0, \quad \forall z \in \overline{\Omega}.$$

再重复情形 1 的证明, 则定理得证. □

定理 2.5 保证凸集的分离, 其中两个凸集是不相交的. 现在考虑两个凸集可以有公共点的情形, 如图 2.3 所示, 引入变分分析处理凸集和非凸集合的集合极性的概念, 只是这里强调所考虑问题的内蕴最优性 / 极性结构; 更多细节可参看文献 [14] 等.

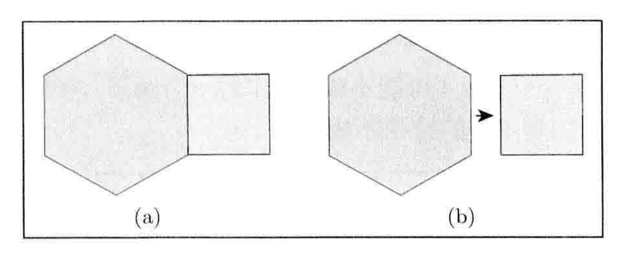


图 2.3　集合的极点系统

定义 2.6　给定两个非空凸集 $\Omega_1, \Omega_2 \subset \mathbb{R}^n$, 称集合系统 $\{\Omega_1, \Omega_2\}$ 是极点系统, 如果对任意的 $\varepsilon > 0$, 存在 $a \in \mathbb{R}^n$ 使得 $\|a\| < \varepsilon$ 且

$$(\Omega_1 - a) \cap \Omega_2 = \varnothing.$$

下面的命题将给出集合极点系统的一个重要例子.

命题 2.7　设 Ω 是 $\mathbb{R}^n$ 的非空凸子集, 设 $\bar{x} \in \Omega$ 是 Ω 的边界点, 则集合 Ω 和 $\{\bar{x}\}$ 构成了一个极点系统.

证明　由于 $\bar{x} \in \operatorname{bd}\Omega$, 对任意数 $\varepsilon > 0$, 有 $\mathbb{B}(\bar{x}; \varepsilon/2) \cap \Omega^c \neq \varnothing$. 所以选取 $b \in \mathbb{B}(\bar{x}; \varepsilon/2) \cap \Omega^c$, 记 $a := b - \bar{x}$, 有 $\|a\| < \varepsilon$ 和 $(\Omega - a) \cap \{\bar{x}\} = \varnothing$. □

下面的结果是凸集的极点系统的分离定理. 它有推广到非凸情形的更一般版本, 被称为极点原理, 是变分分析及其应用的核心[14]. 值得一提的是, 与变分分析的一般情形不同, 在这里对所论问题中的集合不强加任何闭性假设. 这将使我们在本章后面利用变分分析中通常的简单的变分工具, 在没有强加闭性和下半连续性的假设下得到凸集的法向量和凸函数的次梯度的各种运算法则.

定理 2.8　设 Ω_1 和 Ω_2 是 $\mathbb{R}^n$ 的两个非空凸子集, 则 Ω_1 和 Ω_2 构成一个极点系统, 当且仅当存在非零元素 $v \in \mathbb{R}^n$, 在式 (2.2) 的意义下它分离集合 Ω_1 与 Ω_2.

证明　假设 Ω_1 和 Ω_2 构成 $\mathbb{R}^n$ 中凸集的一个极点系统, 则存在 $\mathbb{R}^n$ 中的序列 $\{a_k\}$ 满足 $a_k \to 0$, 且

$$(\Omega_1 - a_k) \cap \Omega_2 = \varnothing, \quad k \in \mathbb{N}.$$

对任意的 $k \in \mathbb{N}$, 应用定理 2.5 于凸集 $\Omega_1 - a_k$ 和 Ω_2, 找到非零元素 $u_k \in \mathbb{R}^n$ 满足

$$\langle u_k, x - a_k \rangle \leqslant \langle u_k, y \rangle, \quad \forall x \in \Omega_1,\ y \in \Omega_2,\ k \in \mathbb{N}.$$

作单位化 $v_k := \dfrac{u_k}{\|u_k\|}$, 则 $\|v_k\| = 1$, 且

$$\langle v_k, x - a_k \rangle \leqslant \langle v_k, y \rangle, \quad \forall x \in \Omega_1, y \in \Omega_2. \tag{2.3}$$

因此可找到 $v \in \mathbb{R}^n$ 且 $\|v\| = 1$ 使得当沿集子序列 $k \to \infty$ 时有 $v_k \to v$. 在式 (2.3) 中取极限, 就证得式 (2.2) 成立.

对于相反的蕴涵关系, 假设存在非零元素 $v \in \mathbb{R}^n$ 在式 (2.2) 的意义下分离集合 Ω_1 与 Ω_2, 则不难看到

$$\left(\Omega_1 - \frac{1}{k}v\right) \cap \Omega_2 = \varnothing, \quad \forall k \in \mathbb{N}.$$

因此 Ω_1 与 Ω_2 构成了一个极点系统. □

2.2 凸集的法向量

本节利用上面得到的分离结果 (特别地, 定理 2.8 中的极点版本), 研究凸集的广义法向量并得到法锥运算的基本法则.

定义 2.9 设 $\Omega \in \mathbb{R}^n$ 是凸集且 $\bar{x} \in \Omega$, Ω 在 $\bar{x}$ 的法锥定义为

$$N(\bar{x}; \Omega) := \{v \in \mathbb{R}^n \mid \langle v, x - \bar{x} \rangle \leqslant 0, \forall x \in \Omega\}. \tag{2.4}$$

按惯例, 置 $N(\bar{x}; \Omega) := \varnothing$, 当 $\bar{x} \notin \Omega$.

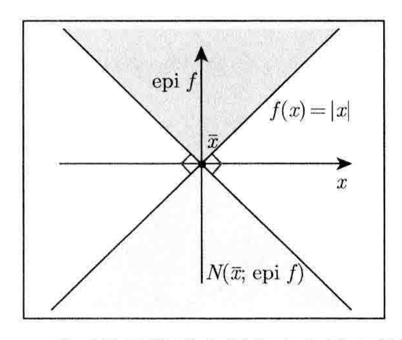


图 2.4 绝对值函数的上图和它在原点的法锥

首先列出上面所定义的法锥的一些基本性质.

命题 2.10 设 $\bar{x} \in \Omega$, Ω 是 $\mathbb{R}^n$ 的凸子集, 则有

(i) $N(\bar{x};\Omega)$ 是含原点的闭凸锥.

(ii) 如果 $\bar{x}\in\operatorname{int}\Omega$, 则 $N(\bar{x};\Omega)=\{0\}$.

证明 为证 (i) 成立, 取 $v_i\in N(\bar{x};\Omega)$, $\lambda_i\geqslant 0$, $i=1,2$. 由定义得 $0\in N(\bar{x};\Omega)$, 且 $\langle v_i,x-\bar{x}\rangle\leqslant 0$, $\forall x\in\Omega$, 蕴涵着

$$\langle\lambda_1v_1+\lambda_2v_2,x-\bar{x}\rangle\leqslant 0,\quad\forall x\in\Omega.$$

因此 $\lambda_1v_1+\lambda_2v_2\in N(\bar{x};\Omega)$, 从而 $N(\bar{x};\Omega)$ 是凸锥. 为验证它是闭的, 取收敛于 v 的序列 $\{v_k\}\subset N(\bar{x};\Omega)$. 当 $k\to\infty$ 时在

$$\langle v_k,x-\bar{x}\rangle\leqslant 0,\quad\forall x\in\Omega$$

中取极限得 $\langle v,x-\bar{x}\rangle\leqslant 0$, $\forall x\in\Omega$, 因此 $v\in N(\bar{x};\Omega)$. 这就完成了 (i) 的证明.

为证 (ii) 成立, 取 $v\in N(\bar{x};\Omega)$, 其中 $\bar{x}\in\operatorname{int}\Omega$, 则可找到 $\delta>0$, 使得 $\bar{x}+\delta\mathbb{B}\subset\Omega$. 因此对任意的 $e\in\mathbb{B}$ 有不等式

$$\langle v,\bar{x}+\delta e-\bar{x}\rangle\leqslant 0$$

成立, 这表明 $\delta\langle v,e\rangle\leqslant 0$, 因此 $\langle v,e\rangle\leqslant 0$. 选取如此小的 $t>0$ 使得 $tv\in\mathbb{B}$, 则 $\langle v,tv\rangle\leqslant 0$, 因此 $t\|v\|^2=0$, 即 $v=0$. □

下面开始建立得到相当简单然而非常有用的运算法则.

命题 2.11 (i) 设 Ω_1 与 Ω_2 分别是 $\mathbb{R}^n$ 与 $\mathbb{R}^p$ 的非空凸子集, 则对 $(\bar{x}_1,\bar{x}_2)\in\Omega_1\times\Omega_2$, 有

$$N((\bar{x}_1,\bar{x}_2);\Omega_1\times\Omega_2)=N(\bar{x}_1;\Omega_1)\times N(\bar{x}_2;\Omega_2).$$

(ii) 设 Ω_1 与 Ω_2 是 $\mathbb{R}^n$ 的凸子集且 $\bar{x}_i\in\Omega_i$, $i=1,2$, 则

$$N(\bar{x}_1+\bar{x}_2;\Omega_1+\Omega_2)=N(\bar{x}_1;\Omega_1)\cap N(\bar{x}_2;\Omega_2).$$

证明 为证 (i) 成立, 取 $(v_1,v_2)\in N((\bar{x}_1,\bar{x}_2);\Omega_1\times\Omega_2)$, 由定义得

$$\langle(v_1,v_2),(x_1,x_2)-(\bar{x}_1,\bar{x}_2)\rangle=\langle v_1,x_1-\bar{x}_1\rangle+\langle v_2,x_2-\bar{x}_2\rangle\leqslant 0,\tag{2.5}$$

对任意的 $(x_1,x_2)\in\Omega_1\times\Omega_2$ 成立. 在式 (2.5) 中置 $x_2:=\bar{x}_2$, 有

$$\langle v_1,x_1-\bar{x}_1\rangle\leqslant 0,\quad\forall x_1\in\Omega_1,$$

这意味着 $v_1\in N(\bar{x}_1;\Omega_1)$. 同理, 得 $v_2\in N(\bar{x}_2;\Omega_2)$, 因此证明了 (1) 中的包含关系 “$\subset$”. 反向的包含关系是显然的.

现证 (ii) 成立. 取 $v\in N(\bar{x}_1+\bar{x}_2;\Omega_1+\Omega_2)$, 由定义得

$$\langle v,x_1+x_2-(\bar{x}_1+\bar{x}_2)\rangle\leqslant 0,\quad\forall x_1\in\Omega_1,x_2\in\Omega_2.$$

在上式中令 $x_1 := \bar{x}_1$, $x_2 := \bar{x}_2$, 则得 $v \in N(\bar{x}_1;\Omega_1) \cap N(\bar{x}_2;\Omega_2)$. (ii) 中相反的包含关系也是不难得到的. □

进一步, 考虑线性映射 $A : \mathbb{R}^n \to \mathbb{R}^p$, 按惯例它对应于一个 $p \times n$ 矩阵. 其伴随映射 $A^* : \mathbb{R}^p \to \mathbb{R}^n$ 定义为

$$\langle Ax, y\rangle = \langle x, A^*y\rangle, \quad \forall\, x \in \mathbb{R}^n,\, y \in \mathbb{R}^p,$$

它对应于矩阵转置. 下面的运算结果描述了线性方程组解集的法向量.

命题 2.12 设 $B : \mathbb{R}^n \to \mathbb{R}^p$ 定义为 $B(x) := Ax + b$, 这里 $A : \mathbb{R}^n \to \mathbb{R}^p$ 是线性映射, $b \in \mathbb{R}^p$. 给定 $c \in \mathbb{R}^p$, 考虑解集

$$\Omega := \{x \in \mathbb{R}^n \mid Bx = c\}.$$

对任意的 $\bar{x} \in \Omega$, 即 $B(\bar{x}) = c$, 有

$$N(\bar{x};\Omega) = \operatorname{im} A^* = \{v \in \mathbb{R}^n \mid v = A^*y,\, y \in \mathbb{R}^p\}. \tag{2.6}$$

证明 取 $v \in N(\bar{x};\Omega)$, 对任意满足 $Ax + b = c$ 的 x 成立, 则 $\langle v, x - \bar{x}\rangle \leqslant 0$. 任取定 $u \in \ker A$, 可以看到 $A(\bar{x} - u) = A\bar{x} = c - b$, 即 $A(\bar{x} - u) + b = c$, 因此对任意的 $u \in \ker A$, $\langle v, u\rangle \geqslant 0$ 成立. 由于 $-u \in \ker A$, 对任意的 $u \in \ker A$ 有 $\langle v, u\rangle = 0$. 用反证法, 假设不存在 $y \in \mathbb{R}^p$ 满足 $v = A^*y$. 因此 $v \notin W := A^*\mathbb{R}^p$, 其中集合 $W \subset \mathbb{R}^n$, 显然是非空、闭凸的, 利用命题 2.1 知有非零元素 $\bar{u} \in \mathbb{R}^n$ 满足

$$\sup\{\langle \bar{u}, w\rangle \mid w \in W\} < \langle \bar{u}, v\rangle,$$

由于 $0 \in W$, 上式蕴涵着 $0 < \langle \bar{u}, v\rangle$. 而且

$$t\langle \bar{u}, A^*y\rangle = \langle \bar{u}, A^*(ty)\rangle < \langle \bar{u}, v\rangle, \quad \forall\, t \in \mathbb{R},\, y \in \mathbb{R}^p.$$

因此, $\langle \bar{u}, A^*y\rangle = 0$, 这就证明了对任意的 $y \in \mathbb{R}^p, \langle A\bar{u}, y\rangle = 0$ 成立, 即 $A\bar{u} = 0$. 因此 $\bar{u} \in \ker A$ 且 $\langle v, \bar{u}\rangle > 0$. 这个矛盾证明了式 (2.6) 中的包含关系“$\subset$”成立.

为证明式 (2.6) 中的相反的包含关系成立, 任取 $v \in \mathbb{R}^n$ 满足对某 $y \in \mathbb{R}^p$, 有 $v = A^*y$ 成立. 对任意的 $x \in \Omega$, 有

$$\langle v, x - \bar{x}\rangle = \langle A^*y, x - \bar{x}\rangle = \langle y, Ax - A\bar{x}\rangle = 0,$$

这意味着 $v \in N(\bar{x};\Omega)$, 从而完成了命题的证明. □

下面的命题是利用凸集 (一般来说非闭的) 的法锥来刻画它们的分离性.

命题 2.13 设 Ω_1 与 Ω_2 是 $\mathbb{R}^n$ 的凸子集且 $\bar{x} \in \Omega_1 \cap \Omega_2$, 则下面的断言是等价的.

(i) $\{\Omega_1, \Omega_2\}$ 是 $\mathbb{R}^n$ 中由两个凸集构成的极点系统.

(ii) 对某个 $v \neq 0$, 集合 Ω_1 和 Ω_2 在式 (2.2) 的意义下是分离的.

(iii) $N(\bar{x};\Omega_1) \cap [-N(\bar{x};\Omega_2)] \neq \{0\}$.

证明 (i) 与 (ii) 的等价性由定理 2.8 可得. 假设 (ii) 满足, 由于 $\bar{x} \in \Omega_2$, 有

$$\sup\{\langle v, x\rangle \mid x \in \Omega_1\} \leqslant \langle v, \bar{x}\rangle,$$

所以, 对任意的 $x \in \Omega_1$, 有 $\langle v, x-\bar{x}\rangle \leqslant 0$, 即 $v \in N(\bar{x};\Omega_1)$.

同理, 可证明包含关系 $-v \in N(\bar{x};\Omega_2)$ 成立, 从而得 (iii) 成立.

现假设 (iii) 满足并设 $0 \neq v \in N(\bar{x};\Omega_1) \cap [-N(\bar{x};\Omega_2)]$, 则式 (2.2) 由法锥的定义直接可得. 事实上, 对任意的 $x \in \Omega_1$ 和 $y \in \Omega_2$, 有

$$\langle v, x-\bar{x}\rangle \leqslant 0, \quad \langle -v, y-\bar{x}\rangle \leqslant 0.$$

上式蕴涵着 $\langle v, x-\bar{x}\rangle \leqslant 0 \leqslant \langle v, y-\bar{x}\rangle$, 因此 $\langle v, x\rangle \leqslant \langle v, y\rangle$. □

下面的推论给出边界点的法锥刻画.

推论 2.14 设 Ω 是 $\mathbb{R}^n$ 的非空凸子集, 设 $\bar{x} \in \Omega$, 则 $\bar{x} \in \operatorname{bd}\Omega$ 当且仅当 Ω 的法锥是非平凡的, 即 $N(\bar{x};\Omega) \neq \{0\}$.

证明 充分性由命题 2.10(ii) 可得. 为证必要性, 取 $\bar{x} \in \operatorname{bd}\Omega$, 考虑集合 $\Omega_1 := \Omega, \Omega_2 := \{\bar{x}\}$, 根据命题 2.7 它们构成了一个极点系统. 注意到 $N(\bar{x};\Omega_2) = \mathbb{R}^n$. 由命题 2.13 即得 $N(\bar{x};\Omega_1) \cap [-N(\bar{x};\Omega_2)] = N(\bar{x};\Omega_1) \neq \{0\}$. □

现在我们已经准备好建立凸集法锥运算的主要结果, 即给出集合交的法向量与交中每个集合的法向量之间的关系. 它的证明也是基于集合极点系统的凸分离, 注意到这种方法允许我们处理一般的凸集, 它们不一定是闭的.

定理 2.15 设 Ω_1 和 Ω_2 是 $\mathbb{R}^n$ 的凸子集且 $\bar{x} \in \Omega_1 \cap \Omega_2$, 则对任意的 $v \in N(\bar{x};\Omega_1 \cap \Omega_2)$, 存在 $\lambda \geqslant 0, v_i \in N(\bar{x};\Omega_i)$, $i=1,2$, 使得

$$\lambda v = v_1 + v_2, \quad (\lambda, v_1) \neq (0,0). \tag{2.7}$$

证明 取 $v \in N(\bar{x};\Omega_1 \cap \Omega_2)$, 由定义得

$$\langle v, x-\bar{x}\rangle \leqslant 0, \quad \forall\, x \in \Omega_1 \cap \Omega_2.$$

记 $\Theta_1 := \Omega_1 \times [0,\infty)$, $\Theta_2 := \{(x,\lambda) \mid x \in \Omega_2, \lambda \leqslant \langle v, x-\bar{x}\rangle\}$.

这两个凸集构成了一个极点系统, 由于显然有 $(\bar{x},0) \in \Theta_1 \cap \Theta_2$, 且

$$\Theta_1 \cap (\Theta_2 - (0,a)) = \varnothing, \quad \forall\, a > 0.$$

根据定理 2.8, 找到 $0 \neq (w,\gamma) \in \mathbb{R}^n \times \mathbb{R}$ 使得

$$\langle w,x\rangle + \lambda_1\gamma \leqslant \langle w,y\rangle + \lambda_2\gamma, \quad \forall (x,\lambda_1) \in \Theta_1, \quad (y,\lambda_2) \in \Theta_2. \tag{2.8}$$

易见 $\gamma \leqslant 0$. 事实上, 假设 $\gamma > 0$, 考虑到当 $k > 0$ 时, $(\bar{x},k) \in \Theta_1$ 且 $(\bar{x},0) \in \Theta_2$, 得

$$\langle w,\bar{x}\rangle + k\gamma \leqslant \langle w,\bar{x}\rangle$$

产生了一个矛盾. 接下来有如下两种可能的情形需考虑.

情形 1 $\gamma = 0$. 此时有 $w \neq 0$ 且

$$\langle w,x\rangle \leqslant \langle w,y\rangle, \quad \forall x \in \Omega_1,\ y \in \Omega_2,$$

则由命题 2.13 可得 $w \in N(\bar{x};\Omega_1) \cap [-N(\bar{x};\Omega_2)]$, 从而对 $\lambda = 0, v_1 = w$ 和 $v_2 = -w$, 式 (2.7) 成立.

情形 2 $\gamma < 0$. 此时令 $\mu := -\gamma > 0$, 利用当 $x \in \Omega_1$ 时 $(x,0) \in \Theta_1$ 和 $(\bar{x},0) \in \Theta_2$, 由式 (2.8) 得

$$\langle w,x\rangle \leqslant \langle w,\bar{x}\rangle, \quad \forall x \in \Omega_1.$$

这蕴涵着 $w \in N(\bar{x};\Omega_1)$ 从而 $\dfrac{w}{\mu} \in N(\bar{x};\Omega_1)$. 进一步, 由于 $(\bar{x},0) \in \Theta_1$, 对任意的 $y \in \Omega_2$ 和 $\alpha := \langle v, y-\bar{x}\rangle$, $(y,\alpha) \in \Theta_2$, 由式 (2.8) 得

$$\langle w,\bar{x}\rangle \leqslant \langle w,y\rangle + \gamma\langle v,y-\bar{x}\rangle, \quad \forall y \in \Omega_2.$$

两边同时除以 γ, 得

$$\left\langle \frac{\omega}{\gamma},\bar{x}\right\rangle \geqslant \left\langle \frac{w}{\gamma},y\right\rangle + \langle v,y-\bar{x}\rangle, \quad \forall y \in \Omega_2.$$

这显然蕴涵着不等式

$$\left\langle \frac{w}{\gamma} + v, y-\bar{x}\right\rangle \leqslant 0, \quad \forall y \in \Omega_2,$$

因此 $\dfrac{w}{\gamma} + v = -\dfrac{w}{\mu} + v \in N(\bar{x};\Omega_2)$. 最后, 令 $v_1 := \dfrac{w}{\mu} \in N(\bar{x};\Omega_1), v_2 := -\dfrac{w}{\mu} + v \in N(\bar{x};\Omega_2)$, 则得 $v = v_1 + v_2$, 这保证当 $\lambda = 1$ 时式 (2.7) 成立. □

注意到表达式 (2.7) 给出了 $\lambda \neq 0$ 时表示集合交的法向量的运算法则. 然而, 当追加下面的基本规范条件时, 根据定理 2.15 这样的交法则是容易得到的. 这个基本规范条件在随后的运算结果中起着至关重要的作用, 参见图 2.5.

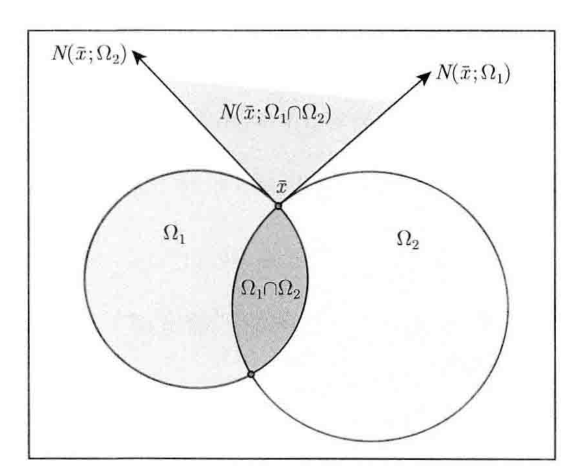


图 2.5 集合交的法锥

推论 2.16 设 Ω_1 和 Ω_2 是 $\mathbb{R}^n$ 的凸子集且 $\bar{x}\in\Omega_1\cap\Omega_2$. 假设基本规范条件 (BQC) 成立:

$$N(\bar{x};\Omega_1)\cap[-N(\bar{x};\Omega_2)]=\{0\}, \tag{2.9}$$

则有法锥交法则:

$$N(\bar{x};\Omega_1\cap\Omega_2)=N(\bar{x};\Omega_1)+N(\bar{x};\Omega_2). \tag{2.10}$$

证明 对任意的 $v\in N(\bar{x};\Omega_1\cap\Omega_2)$, 根据定理 2.15 找到数 $\lambda\geqslant 0$ 和向量 $v_i\in N(\bar{x};\Omega_i)$, $i=1,2$, 使得式 (2.7) 中的条件成立. 若在式 (2.7) 中假设 $\lambda=0$, 则有 $0\neq v_1=-v_2\in N(\bar{x};\Omega_1)\cap[-N(\bar{x};\Omega_2)]$, 与 BQC (2.9) 矛盾. 因此 $\lambda>0$ 且 $v=v_1/\lambda+v_2/\lambda\in N(\bar{x};\Omega_1)+N(\bar{x};\Omega_2)$. 因此, 式 (2.10) 中的包含关系“$\subset$”满足. 式 (2.10) 中相反的包含关系利用法锥的定义易证. 事实上, 任取 $v\in N(\bar{x};\Omega_1)+N(\bar{x};\Omega_2)$ 且 $v=v_1+v_2$, 其中 $v_i\in N(\bar{x};\Omega_i)$, $i=1,2$, 则对任意的 $x\in\Omega_1\cap\Omega_2$, 有

$$\langle v,x-\bar{x}\rangle=\langle v_1+v_2,x-\bar{x}\rangle=\langle v_1,x-\bar{x}\rangle+\langle v_2,x-\bar{x}\rangle\leqslant 0,$$

这蕴涵着 $v\in N(\bar{x};\Omega_1\cap\Omega_2)$, 从而完成了推论的证明. □

下面的例子表明如果 BQC(2.9) 不成立, 那么交法则 (2.10) 一般来说是不成立的.

例 2.17 设 $\Omega_1:=\{(x,\lambda)\in\mathbb{R}^2\mid\lambda\geqslant x^2\},\Omega_2:=\{(x,\lambda)\in\mathbb{R}^2\mid\lambda\leqslant -x^2\}$, 则对 $\bar{x}=(0,0)$, 有 $N(\bar{x};\Omega_1\cap\Omega_2)=\mathbb{R}^2$, 而 $N(\bar{x};\Omega_1)+N(\bar{x};\Omega_2)=\{0\}\times\mathbb{R}$.

接下来给出一个易于验证的条件保证 BQC(2.9) 成立. 针对相对内部的更一般结果在练习 2.13 及其答案中给出.

推论 2.18 设 $\Omega_1,\Omega_2\subset\mathbb{R}^n$ 是凸集满足 $\operatorname{int}\Omega_1\cap\Omega_2\neq\varnothing$ 或 $\operatorname{int}\Omega_2\cap\Omega_1\neq\varnothing$, 则对任意的 $\bar{x}\in\Omega_1\cap\Omega_2$, BQC(2.9) 和交法则式 (2.10) 均满足.

证明 确定起见, 考虑当 $\operatorname{int}\Omega_1 \cap \Omega_2 \neq \varnothing$ 的情形. 取 $\bar{x} \in \Omega_1 \cap \Omega_2$, 只需证明 BQC(2.9) 成立, 从而根据推论 2.16 保证交法则式 (2.10) 成立. 用反证法, 假设

$$N(\bar{x};\Omega_1) \cap [-N(\bar{x};\Omega_2)] \neq \{0\},$$

则根据命题 2.13 找到一个非零元素 $v \in \mathbb{R}^n$ 像式 (2.2) 那样分离集合 Ω_1 和 Ω_2. 现取 $\bar{u} \in \operatorname{int}\Omega_1 \cap \Omega_2$, 存在数 $\delta > 0$ 使得 $\bar{u} + \delta\mathbb{B} \subset \Omega_1$. 根据式 (2.2) 确保

$$\langle v, x\rangle \leqslant \langle v, \bar{u}\rangle, \quad \forall\, x \in \bar{u} + \delta\mathbb{B}.$$

这进一步蕴涵着不等式

$$\langle v, \bar{u}\rangle + \delta\langle v, e\rangle \leqslant \langle v, \bar{u}\rangle, \quad \forall\, e \in \mathbb{B},$$

从而对任意的 $e \in \mathbb{B}$, 有 $\langle v, e\rangle \leqslant 0$. 最后选取 $e := \dfrac{v}{\|v\|}$, 得到一个矛盾, 因此推论得证. □

利用数学归纳法可得到对有限多个集合的推论 2.16(以及类似地, 推论 2.18) 的相应结果.

推论 2.19 设 $\Omega_i \subset \mathbb{R}^n$, $i = 1, \cdots, m$ 是凸集, 且 $\bar{x} \in \bigcap_{i=1}^m \Omega_i$. 假设下面的规范条件成立:

$$\left[\sum_{i=1}^m v_i = 0, v_i \in N(\bar{x};\Omega_i)\right] \Rightarrow [v_i = 0, i = 1, \cdots, m],$$

则有交公式

$$N\left(\bar{x};\bigcap_{i=1}^m \Omega_i\right) = \sum_{i=1}^m N(\bar{x};\Omega_i).$$

2.3 凸函数的 Lipschitz 连续性

本节研究凸函数 Lipschitz 连续性的基本概念. 这个概念可被认为是“线性率连续性”, 且在凸分析、变分分析及其应用的许多方面是非常重要的.

定义 2.20 函数 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 在集合 $\Omega \subset \operatorname{dom} f$ 上是 Lipschitz 连续的, 如果存在常数 $\ell \geqslant 0$ 使得

$$|f(x) - f(y)| \leqslant \ell\, \| x - y \|, \quad \forall\, x, y \in \Omega. \tag{2.11}$$

称 f 在 $\bar{x} \in \operatorname{dom} f$ 附近是局部 Lipschitz 连续的, 如果存在常数 $\ell \geqslant 0$ 和 $\delta > 0$ 使得对 $\Omega = \mathbb{B}(\bar{x};\delta)$ 式 (2.11) 成立.

这里给出一般凸函数几个可验证的条件确保 Lipschitz 连续性并且刻画它的局部化版本. 局部 Lipschitz 连续性最有效的刻画之一, 对于非凸函数也成立, 是通过下面奇异 (或水平) 次微分的概念给出的, 它在变分分析的框架中才被正确认识, 而在凸分析中被大多人忽视了, 参见文献 [14] 和 [27]. 注意到, 与在 2.4 节中定义的凸分析 (通常的) 次微分不同, 奇异次微分不简化为可微函数的经典导数.

定义 2.21　设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸函数, $\bar{x}\in\operatorname{dom}f$, 函数 f 在 $\bar{x}$ 的奇异次微分定义为

$$\partial^\infty f(\bar{x}):=\{v\in\mathbb{R}^n\mid(v,0)\in N((\bar{x},f(\bar{x}));\operatorname{epi}f)\}. \tag{2.12}$$

我们用下面的例子来说明基于定义式 (2.12) 的 $\partial^\infty f(\bar{x})$ 的计算.

例 2.22　考虑函数

$$f(x):=\begin{cases}0, & |x|\leqslant 1,\\ \infty, & \text{其他},\end{cases}$$

则 $\operatorname{epi}f=[-1,1]\times[0,\infty)$, 对 $\bar{x}=1$, 有

$$N((\bar{x},f(\bar{x}));\operatorname{epi}f)=[0,\infty)\times(-\infty,0],$$

这表明 $\partial^\infty f(\bar{x})=[0,\infty)$, 参见图 2.6.

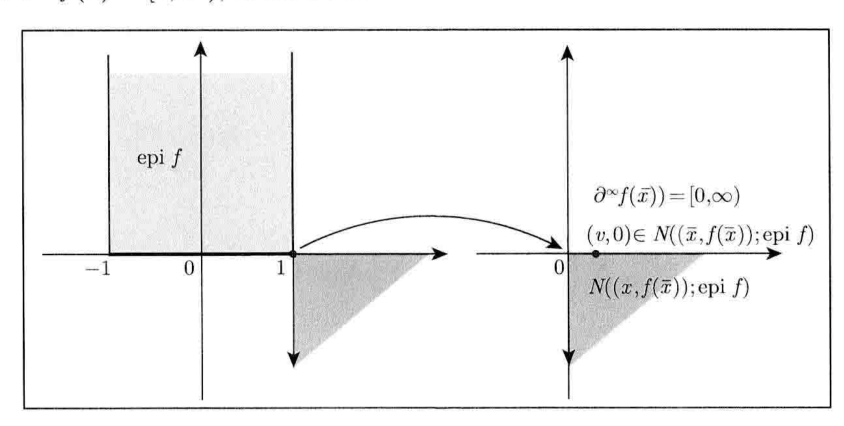


图 2.6　奇异次梯度

下面的命题大大地简化了 $\partial^\infty f(\bar{x})$ 的计算.

命题 2.23　对凸函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 且 $\bar{x}\in\operatorname{dom}f$, 有

$$\partial^\infty f(\bar{x})=N(\bar{x};\operatorname{dom}f).$$

证明　任取 $v\in\partial^\infty f(\bar{x})$ 且 $x\in\operatorname{dom}f$. 注意到 $(x,f(x))\in\operatorname{epi}f$, 则利用式 (2.12) 和法锥结构式 (2.4), 有

$$\langle v,x-\bar{x}\rangle=\langle v,x-\bar{x}\rangle+0(f(x)-f(\bar{x}))\leqslant 0,$$

这表明 $v \in N(\bar{x}; \operatorname{dom} f)$. 反之, 假设 $v \in N(\bar{x}; \operatorname{dom} f)$, 任取 $(x, \lambda) \in \operatorname{epi} f$, 则有 $f(x) \leqslant \lambda$, 因此 $x \in \operatorname{dom} f$. 于是

$$\langle v, x-\bar{x}\rangle + 0(\lambda - f(\bar{x})) = \langle v, x-\bar{x}\rangle \leqslant 0,$$

这蕴涵着 $(v, 0) \in N((\bar{x}, f(\bar{x})); \operatorname{epi} f)$, 即 $v \in \partial^{\infty} f(\bar{x})$. □

为继续研究凸函数的 Lipschitz 连续性, 需要下面两个引理, 它们本身也有独立的意义.

引理 2.24 设 $\{e_i \mid i=1,\cdots,n\}$ 是 $\mathbb{R}^n$ 的标准正交基. 记

$$A := \{\bar{x} \pm \varepsilon e_i \mid i=1,\cdots,n\}, \quad \varepsilon > 0,$$

则下面的性质成立:

(i) 对 $|\gamma| \leqslant \varepsilon, i=1,\cdots,n, \bar{x}+\gamma e_i \in \operatorname{co} A$.

(ii) $\mathbb{B}(\bar{x}; \varepsilon/n) \subset \operatorname{co} A$.

证明 (i) 对 $|\gamma| \leqslant \varepsilon$, 找到 $t \in [0,1]$ 满足 $\gamma = t(-\varepsilon)+(1-t)\varepsilon$, 则由 $\bar{x} \pm \varepsilon e_i \in A$ 得

$$\bar{x}+\gamma e_i = t(\bar{x}-\varepsilon e_i) + (1-t)(\bar{x}+\varepsilon e_i) \in \operatorname{co} A.$$

(ii) 对 $x \in \mathbb{B}(\bar{x}; \varepsilon/n)$, 有 $x = \bar{x} + (\varepsilon/n)u$, 其中 $\|u\| \leqslant 1$. 通过 $\{e_i\}$ 把 u 表示为

$$u = \sum_{i=1}^{n} \lambda_i e_i,$$

其中对任意的 i, $|\lambda_i| \leqslant \sqrt{\sum_{i=1}^{n} \lambda_i^2} = \|u\| \leqslant 1$. 从而有

$$x = \bar{x} + \frac{\varepsilon}{n}u = \bar{x} + \sum_{i=1}^{n} \frac{\varepsilon\lambda_i}{n} e_i = \sum_{i=1}^{n} \frac{1}{n}(\bar{x}+\varepsilon\lambda_i e_i).$$

记 $\gamma_i := \varepsilon\lambda_i$, 则 $|\gamma_i| \leqslant \varepsilon$. 由 (i) 得 $\bar{x}+\varepsilon\lambda_i e_i = \bar{x}+\gamma_i e_i \in \operatorname{co} A$, 由于 x 是 $\operatorname{co} A$ 中元素的凸组合, 因此 $x \in \operatorname{co} A$. □

引理 2.25 如果对某元素 $\bar{x} \in \operatorname{dom} f$, 数 $\delta > 0$, 凸函数 $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ 在 $\mathbb{B}(\bar{x}; \delta)$ 上是上有界的, 则 f 在 $\mathbb{B}(\bar{x}; \delta)$ 上是有界的.

证明 记 $m := f(\bar{x})$, 并取 $M > 0$ 使得对任意的 $x \in \mathbb{B}(\bar{x}; \delta)$ 有 $f(x) \leqslant M$. 任取 $u \in \mathbb{B}(\bar{x}; \delta)$, 考虑元素 $x := 2\bar{x} - u$, 则 $x \in \mathbb{B}(\bar{x}; \delta)$ 且

$$m = f(\bar{x}) = f\left(\frac{x+u}{2}\right) \leqslant \frac{1}{2}f(x) + \frac{1}{2}f(u),$$

这表明 $f(u) \geqslant 2m - f(x) \geqslant 2m - M$, 因此 f 在 $\mathbb{B}(\bar{x};\delta)$ 上是有界的. □

下面的重要结果表明凸函数的局部有界性蕴涵它的局部 Lipschitz 连续性.

定理 2.26 设 $f:\mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸的且 $\bar{x} \in \operatorname{dom} f$. 如果对某 $\delta > 0$, f 在 $\mathbb{B}(\bar{x};\delta)$ 上是上有界的, 则 f 在 $\mathbb{B}(\bar{x};\delta/2)$ 上是 Lipschitz 连续的.

证明 取定 $x, y \in \mathbb{B}(\bar{x};\delta/2)$ 且 $x \neq y$, 考虑元素

$$u := x + \frac{\delta}{2\|x-y\|}(x-y).$$

由于 $e := \dfrac{x-y}{\|x-y\|} \in \mathbb{B}$, 因此有 $u = x + \dfrac{\delta}{2}e \in \bar{x} + \dfrac{\delta}{2}\mathbb{B} + \dfrac{\delta}{2}\mathbb{B} \subset \bar{x} + \delta\mathbb{B}$. 进一步, 记 $\alpha := \dfrac{\delta}{2\|x-y\|}$, 则 $u = x + \alpha(x-y)$, 因此

$$x = \frac{1}{\alpha+1}u + \frac{\alpha}{\alpha+1}y.$$

由 f 的凸性得

$$f(x) \leqslant \frac{1}{\alpha+1}f(u) + \frac{\alpha}{\alpha+1}f(y),$$

这反过来蕴涵着不等式

$$f(x) - f(y) \leqslant \frac{1}{\alpha+1}(f(u) - f(y)) \leqslant 2M\frac{1}{\alpha+1} = 2M\frac{2\|x-y\|}{\delta + 2\|x-y\|} \leqslant \frac{4M}{\delta}\|x-y\|,$$

其中 $M := \sup\{|f(x)| \mid x \in \mathbb{B}(\bar{x};\delta)\}$ (根据引理 2.25, $M < \infty$). 在上式中互换 x 与 y, 则得

$$|f(x) - f(y)| \leqslant \frac{4M}{\delta}\|x-y\|,$$

这就证明了 f 在 $\mathbb{B}(\bar{x};\delta/2)$ 上的 Lipschitz 连续性. □

下面的两个推论确保凸函数在某些特定集合上的 Lipschitz 连续性自动成立.

推论 2.27 任意增广实值凸函数 $f:\mathbb{R}^n \to \overline{\mathbb{R}}$ 在定义域的内部 $\operatorname{int}(\operatorname{dom} f)$ 上是局部 Lipschitz 连续的.

证明 取 $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$, 选取 $\varepsilon > 0$ 使得对任意的 i 有 $\bar{x} \pm \varepsilon e_i \in \operatorname{dom} f$. 考虑引理 2.24 中的集合 A, 由这个引理的断言 (ii) 得 $\mathbb{B}(\bar{x};\varepsilon/n) \subset \operatorname{co} A$. 在有限集合 A 上, 记 $M := \max\{f(a) \mid a \in A\} < \infty$. 对任意的 $x \in \mathbb{B}(\bar{x};\varepsilon/n)$, 利用表示

$$x = \sum_{i=1}^m \lambda_i a_i, \quad 其中 \quad \lambda_i \geqslant 0, \quad \sum_{i=1}^m \lambda_i = 1, \quad a_i \in A,$$

得

$$f(x) \leqslant \sum_{i=1}^m \lambda_i f(a_i) \leqslant \sum_{i=1}^m \lambda_i M = M,$$

因此 f 在 $\mathbb{B}(\bar{x};\varepsilon/n)$ 上是上有界的. 于是由定理 2.26 知 f 在 $\mathbb{B}(\bar{x};\varepsilon/2n)$ 上是 Lipschitz 连续的, 从而它在 $\operatorname{int}(\operatorname{dom} f)$ 上是局部 Lipschitz 连续的. □

由推论 2.27 立即可得 $\mathbb{R}^n$ 上的任何有限凸函数在整个空间上总是局部 Lipschitz 连续的.

推论 2.28 如果 $f:\mathbb{R}^n\to\mathbb{R}$ 是凸的, 则它在 $\mathbb{R}^n$ 上是局部 Lipschitz 连续的.

下面给出了增广实值凸函数局部 Lipschitz 连续性的几个刻画.

定理 2.29 设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸的且 $\bar{x}\in\operatorname{dom} f$, 则下列条件是等价的:

(i) f 在 $\bar{x}$ 处是连续的.

(ii) $\bar{x}\in\operatorname{int}(\operatorname{dom} f)$.

(iii) f 在 $\bar{x}$ 附近是局部 Lipschitz 连续的.

(iv) $\partial^\infty f(x)=\{0\}$.

证明 为证 (i)⇒(ii), 根据 f 在 $\bar{x}$ 的连续性, 对任意的 $\varepsilon>0$, 找到 $\delta>0$ 使得

$$|f(x)-f(\bar{x})|<\varepsilon,\quad \forall\, x\in\mathbb{B}(\bar{x};\delta).$$

由于 $f(\bar{x})<\infty$, 这蕴涵着 $\mathbb{B}(\bar{x};\delta)\subset\operatorname{dom} f$, 所以 $\bar{x}\in\operatorname{int}(\operatorname{dom} f)$.

蕴涵关系 (ii)⇒(iii) 由推论 2.27 直接可得, 而 (iii)⇒(i) 是显然的. 因此, 条件 (i)~(iii) 是等价的. 接下来为证 (ii) ⇒ (iv), 取 $\bar{x}\in\operatorname{int}(\operatorname{dom} f)$, 则 $N(\bar{x};\operatorname{dom} f)=\{0\}$. 由命题 2.23 得 $\partial^\infty f(\bar{x})=\{0\}$.

最后, 为证 (iv)⇒(ii), 用反证法, 假设 $\bar{x}\notin\operatorname{int}(\operatorname{dom} f)$, 则 $\bar{x}\in\operatorname{bd}(\operatorname{dom} f)$, 由推论 2.14 得 $N(\bar{x};\operatorname{dom} f)\neq\{0\}$, 根据命题 2.23 这与条件 $\partial^\infty f(\bar{x})=\{0\}$ 矛盾. □

2.4 凸函数的次梯度

本节定义并开始研究凸增广实值函数的次微分(次梯度的集合)的概念. 广义导数的这个概念是分析中最基本的概念之一, 具有众多重要的应用. 次微分概念背后的革命性的思想是其集值性, 这表征了函数在单个参考点处的非光滑性, 也将其与其他类型的广义导数概念区别开来. 次微分的集值性在建立它的分析法则及应用中虽然提供了许多有利条件和灵活性, 但是它也造成了某些困难. 次微分分析的主要技术围绕集合的凸分离展开.

定义 2.30 设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸函数, 且 $\bar{x}\in\operatorname{dom} f$. 元素 $v\in\mathbb{R}^n$ 称为 f 在 $\bar{x}$ 的一个次梯度, 如果

$$\langle v,x-\bar{x}\rangle\leqslant f(x)-f(\bar{x}),\quad \forall x\in\mathbb{R}^n. \tag{2.13}$$

f 在 $\bar{x}$ 的所有次梯度的全体称为函数 f 在该点的次微分, 记为 $\partial f(\bar{x})$.

注意到在次梯度定义式 (2.13) 中取 $x \in \operatorname{dom} f$ 就可以了. 也注意到在更一般的条件下 $\partial f(\bar{x}) \neq \varnothing$. 例如, 参见命题 2.47 以及文献[26]中给出该方向的其他结果.

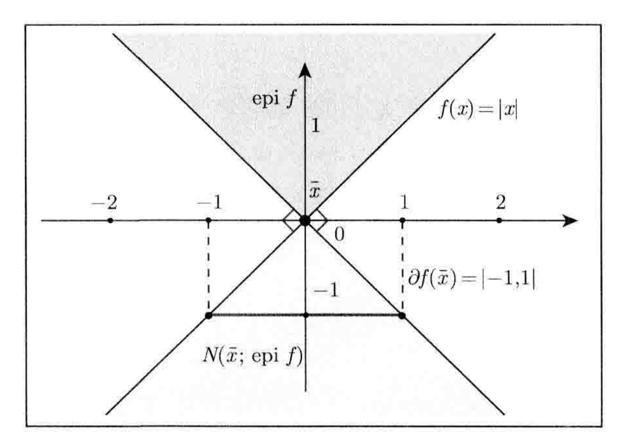


图 2.7　绝对值函数在原点的次微分

下面的命题指明次微分 $\partial f(\bar{x})$ 可从几何上通过 f 的上图在 $(\bar{x}, f(\bar{x}))$ 处的法锥等价地定义.

命题 2.31　对任意凸函数 $f:\mathbb{R}^n \to \overline{\mathbb{R}}$, 任意 $\bar{x} \in \operatorname{dom} f$, 有

$$\partial f(\bar{x}) = \{v \in \mathbb{R}^n \mid (v,-1) \in N((\bar{x}, f(x)); \operatorname{epi} f)\}. \tag{2.14}$$

证明　取 $v \in \partial f(\bar{x}), (x,\lambda) \in \operatorname{epi} f$. 由于 $\lambda \geqslant f(x)$, 由式 (2.13) 可得

$$\begin{aligned}\langle (v,-1),(x,\lambda)-(\bar{x},f(\bar{x}))\rangle &= \langle v, x-\bar{x}\rangle - (\lambda - f(\bar{x}))\\ &\leqslant \langle v, x-\bar{x}\rangle - (f(x)-f(\bar{x})) \leqslant 0.\end{aligned}$$

这就蕴涵着 $(v,-1) \in N((\bar{x}, f(\bar{x})); \operatorname{epi} f)$.

为证式 (2.14) 中相反的包含关系成立, 任取元素 $v \in \mathbb{R}^n$ 使得 $(v,-1) \in N((\bar{x}, f(\bar{x})); \operatorname{epi} f)$. 对任意的 $x \in \operatorname{dom} f$, 有 $(x, f(x)) \in \operatorname{epi} f$. 因此

$$\langle v, x-\bar{x}\rangle - (f(x)-f(\bar{x})) = \langle (v,-1),(x,f(x))-(\bar{x},f(\bar{x}))\rangle \leqslant 0,$$

这就表明 $v \in \partial f(\bar{x})$, 证毕.　□

与利用式 (2.14) 中的法锥表示次微分相对应, 有下面的简单方法通过增广实指示函数的次微分(以及奇异次微分式 (2.12))来表达集合法锥.

例 2.32　给定非空凸集 $\Omega \subset \mathbb{R}^n$, 定义它的指示函数为

$$f(x) = \delta(x;\Omega) := \begin{cases} 0, & x \in \Omega, \\ \infty, & \text{其他}, \end{cases} \tag{2.15}$$

则 $\text{epi}\, f = \Omega \times [0, \infty)$, 根据命题 2.11(i) 得

$$N((\bar{x}, f(\bar{x})); \text{epi}\, f) = N(\bar{x}; \Omega) \times (-\infty, 0],$$

这就蕴涵着

$$\partial f(\bar{x}) = \partial^\infty f(\bar{x}) = N(\bar{x}; \Omega). \tag{2.16}$$

凸函数显著的特征是一个凸函数的每个局部极小值点也给出这个函数的全局/绝对极小值.

命题 2.33 设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸的且 $\bar{x} \in \text{dom}\, f$ 是 f 的一个局部极小值点, 则 f 在该点达到它的全局极小值.

证明 由于 $\bar{x}$ 是 f 的局部极小值点, 所以存在 $\delta > 0$ 使得

$$f(u) \geqslant f(\bar{x}), \quad \forall\, u \in \mathbb{B}(\bar{x}; \delta).$$

取定 $x \in \mathbb{R}^n$, 构造序列 $x_k := (1 - k^{-1})\bar{x} + k^{-1}x, k \in \mathbb{N}$. 因此当 k 充分大时, 有 $x_k \in \mathbb{B}(\bar{x}; \delta)$. 由 f 的凸性得

$$f(\bar{x}) \leqslant f(x_k) \leqslant (1 - k^{-1})f(\bar{x}) + k^{-1}f(x),$$

这蕴涵着 $k^{-1}f(\bar{x}) \leqslant k^{-1}f(x)$, 因此对任意的 $x \in \mathbb{R}^n, f(\bar{x}) \leqslant f(x)$. □

接下来回顾 $\mathbb{R}^n$ 上函数导数的经典概念, 除非另作声明, 全书在 Fréchet 意义下理解导数概念; 参见 3.1 节.

定义 2.34 称 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 在 $\bar{x} \in \text{int}(\text{dom}\, f)$ 处是 (Fréchet) 可微的, 如果存在一个元素 $v \in \mathbb{R}^n$, 使得

$$\lim_{x \to \bar{x}} \frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} = 0.$$

此时元素 v 是唯一定义的且记为 $\nabla f(\bar{x}) := v$.

由经典分析的 Fermat 驻点法则知, 如果 $f : \mathbb{R}^n \to \mathbb{R}$ 在 $\bar{x}$ 处是可微的且在该点达到它的局部极小值, 则 $\nabla f(\bar{x}) = 0$. 对一般凸函数, 下面的命题给出了这个结果的次微分版本.

命题 2.35 设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸的, $\bar{x} \in \text{dom}\, f$, 则 f 在 $\bar{x}$ 处达到它的局部/全局极小值当且仅当 $0 \in \partial f(\bar{x})$.

证明 假设 f 在 $\bar{x}$ 处达到它的全局极小值, 则

$$f(\bar{x}) \leqslant f(x), \quad \forall x \in \mathbb{R}^n.$$

上式可改写为

$$0 = \langle 0, x - \bar{x} \rangle \leqslant f(x) - f(\bar{x}), \quad \forall x \in \mathbb{R}^n.$$

由次微分的定义这等价于 $0 \in \partial f(\bar{x})$. □

现在我们证明对可微函数来说次微分 (2.13) 实际上是单点集, 即其简化为在参考点处经典的导数/梯度, 这也澄清了在凸函数情形下可微性的概念.

命题 2.36 设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸的, 且在 $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$ 处可微, 则有 $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$ 和

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \leqslant f(x) - f(\bar{x}), \quad \forall x \in \mathbb{R}^n. \tag{2.17}$$

证明 由 f 在 $\bar{x}$ 处的可微性知对任意的 $\varepsilon > 0$, 存在 $\delta > 0$, 对任意的 $\|x - \bar{x}\| < \delta$, 满足

$$-\varepsilon\|x - \bar{x}\| \leqslant f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle \leqslant \varepsilon\|x - \bar{x}\|, \tag{2.18}$$

进一步考虑凸函数

$$\varphi(x) := f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \varepsilon\|x - \bar{x}\|, \quad x \in \mathbb{R}^n,$$

注意到对任意的 $x \in \mathbb{B}(\bar{x}; \delta)$ 有 $\varphi(x) \geqslant \varphi(\bar{x}) = 0$. φ 的凸性保证对任意的 $x \in \mathbb{R}^n, \varphi(x) \geqslant \varphi(\bar{x})$. 因此

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \leqslant f(x) - f(\bar{x}) + \varepsilon\|x - \bar{x}\|, \quad \forall x \in \mathbb{R}^n,$$

令 $\varepsilon \downarrow 0$, 则得式 (2.17).

由式 (2.17) 得 $\nabla f(\bar{x}) \in \partial f(\bar{x})$. 现取 $v \in \partial f(\bar{x})$, 得

$$\langle v, x - \bar{x} \rangle \leqslant f(x) - f(\bar{x}),$$

则由式 (2.18) 的第二个不等式有

$$\langle v - \nabla f(\bar{x}), x - \bar{x} \rangle \leqslant \varepsilon\|x - \bar{x}\|, \text{ 对任意的} \|x - \bar{x}\| < \delta \text{成立}.$$

所以 $\|v - \nabla f(\bar{x})\| \leqslant \varepsilon$. 由 $\varepsilon > 0$ 的任意性得 $v = \nabla f(\bar{x})$. 因此 $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$. □

推论 2.37 设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是严格凸函数, 则 f 在 $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$ 的可微性蕴涵着严格不等式

$$\langle \nabla f(\bar{x}), x - \bar{x}, \rangle < f(x) - f(\bar{x}), \quad \text{对任意的 } x \neq \bar{x} \text{成立}. \tag{2.19}$$

证明 由于 f 是凸的, 由命题 2.36 得

$$\langle \nabla f(\bar{x}), u - \bar{x} \rangle \leqslant f(u) - f(\bar{x}), \quad \forall u \in \mathbb{R}^n.$$

取定 $x \neq \bar{x}$, 令 $u := (x + \bar{x})/2$. 由上式得

$$\left\langle \nabla f(x), \frac{x + \bar{x}}{2} - \bar{x} \right\rangle \leqslant f\left(\frac{x + \bar{x}}{2}\right) - f(\bar{x}) < \frac{1}{2}f(x) + \frac{1}{2}f(\bar{x}) - f(\bar{x}).$$

因此得严格不等式 (2.19). □

非光滑凸函数的一个简单但重要的例子是 $\mathbb{R}^n$ 上的范数函数, 下面给出其次微分的计算.

例 2.38 设 $p(x) := \|x\|$ 是 $\mathbb{R}^n$ 上的 Euclid 范数, 则有

$$\partial p(x) = \begin{cases} \mathbb{B}, & x = 0, \\ \left\{ \dfrac{x}{\|x\|} \right\}, & \text{其他}. \end{cases}$$

为证明上式成立, 首先注意到 Euclid 范数函数p在任何非零点处是可微的且$\nabla p(x) = \dfrac{x}{\|x\|}$, $x \neq 0$. 余下计算它在 $x = 0$ 处的次微分. 为此, 根据定义式 (2.13) 得 $v \in \partial p(0)$ 当且仅当

$$\langle v, x \rangle = \langle v, x - 0 \rangle \leqslant p(x) - p(0) = \|x\|, \quad \forall\, x \in \mathbb{R}^n.$$

令 $x = v$, 有 $\langle v, v \rangle \leqslant \|v\|$, 这蕴涵着 $\|v\| \leqslant 1$, 即 $v \in \mathbb{B}$. 现取 $v \in \mathbb{B}$, 由 Cauchy-Schwarz 不等式可推出

$$\langle v, x - 0 \rangle = \langle v, x \rangle \leqslant \|v\| \cdot \|x\| \leqslant \|x\| = p(x) - p(0), \quad \forall x \in \mathbb{R}^n,$$

因此 $v \in \partial p(0)$, 这表明 $\partial p(0) = \mathbb{B}$.

接下来的定理计算了距离函数的次微分.

定理 2.39 设 Ω 是 $\mathbb{R}^n$ 的非空闭凸子集, 则有

$$\partial d(\bar{x}; \Omega) = \begin{cases} N(\bar{x}; \Omega) \cap \mathbb{B}, & \bar{x} \in \Omega, \\ \left\{ \dfrac{\bar{x} - \Pi(\bar{x}; \Omega)}{d(\bar{x}; \Omega)} \right\}, & \text{其他}. \end{cases}$$

证明 考虑当 $\bar{x} \in \Omega$ 的情形, 并任取元素 $w \in \partial d(\bar{x}; \Omega)$. 由次微分的定义得

$$\langle w, x - \bar{x} \rangle \leqslant d(x; \Omega) - d(\bar{x}; \Omega) = d(x; \Omega), \quad \forall x \in \mathbb{R}^n. \tag{2.20}$$

由于距离函数 $d(\cdot; \Omega)$ 是具常数 $\ell = 1$ Lipschitz 连续的, 则有

$$\langle w, x - \bar{x} \rangle \leqslant \|x - \bar{x}\|, \quad \forall x \in \mathbb{R}^n.$$

这蕴涵着 $\|w\| \leqslant 1$, 即 $w \in \mathbb{B}$. 由式 (2.20) 又得

$$\langle w, x - \bar{x} \rangle \leqslant 0, \quad \forall\, x \in \Omega.$$

因此 $w \in N(\bar{x}; \Omega)$, 这就证明了 $w \in N(\bar{x}; \Omega) \cap \mathbb{B}$.

为证反向的包含关系, 任取 $w \in N(\bar{x};\Omega) \cap \mathbb{B}$, 则 $\|w\| \leqslant 1$ 且

$$\langle w, u-\bar{x}\rangle \leqslant 0, \quad \forall\, u \in \Omega.$$

因此, 对任意的 $x \in \mathbb{R}^n$, $u \in \Omega$, 有

$$\begin{aligned}\langle w, x-\bar{x}\rangle &= \langle w, x-u+u-\bar{x}\rangle = \langle w, x-u\rangle + \langle w, u-\bar{x}\rangle \\ &\leqslant \langle w, x-u\rangle \leqslant \|w\| \cdot \|x-u \,| \leqslant \|x-u\|.\end{aligned}$$

由于 u 在 Ω 中是任取的, 由上式可得 $\langle w, x-\bar{x}\rangle \leqslant d(x;\Omega) = d(x;\Omega) - d(\bar{x};\Omega)$, 因此 $w \in \partial d(\bar{x};\Omega)$.

现假设 $\bar{x} \notin \Omega$, 取 $\bar{z} := \Pi(\bar{x};\Omega) \in \Omega$, 任取 $w \in \partial d(\bar{x};\Omega)$, 则

$$\begin{aligned}\langle w, x-\bar{x}\rangle &\leqslant d(x;\Omega) - d(\bar{x};\Omega) = d(x;\Omega) - \|\bar{x}-\bar{z}\| \\ &\leqslant \|x-\bar{z}\| - \|\bar{x}-\bar{z}\|, \quad \forall\, x \in \mathbb{R}^n.\end{aligned}$$

记 $p(x) := \|x-\bar{z}\|$, 有

$$\langle w, x-\bar{x}\rangle \leqslant p(x) - p(\bar{x}), \quad \forall\, x \in \mathbb{R}^n,$$

这保证 $w \in \partial p(\bar{x}) = \left\{\dfrac{\bar{x}-\bar{z}}{\|\bar{x}-\bar{z}\|}\right\}$. 现证 $w = \dfrac{\bar{x}-\bar{z}}{\|\bar{x}-\bar{z}\|}$ 是 $d(\cdot;\Omega)$ 在 $\bar{x}$ 处的一个次梯度. 事实上, 对任意的 $x \in \mathbb{R}^n$, 记 $p_x := \Pi(x;\Omega)$, 则有

$$\begin{aligned}\langle w, x-\bar{x}\rangle &= \langle w, x-\bar{z}\rangle + \langle w, \bar{z}-\bar{x}\rangle = \langle w, x-\bar{z}\rangle - \|\bar{x}-\bar{z}\| \\ &= \langle w, x-p_x\rangle + \langle w, p_x-\bar{z}\rangle - \|\bar{x}-\bar{z}\|.\end{aligned}$$

由于 $\bar{z} = \Pi(\bar{x};\Omega)$, 由命题 1.78 得 $\langle \bar{x}-\bar{z}, p_x-\bar{z}\rangle \leqslant 0$, 因此有 $\langle w, p_x-\bar{z}\rangle \leqslant 0$. 利用事实 $\|w\| = 1$ 和 Cauchy-Schwarz 不等式得

$$\begin{aligned}\langle w, x-\bar{x}\rangle &= \langle w, x-p_x\rangle + \langle w, p_x-\bar{z}\rangle - \|\bar{x}-\bar{z}\| \\ &\leqslant \|w\| \cdot \|x-p_x\| - \|\bar{x}-\bar{z}\| = \|x-p_x\| - \|\bar{x}-\bar{z}\| \\ &= d(x;\Omega) - d(\bar{x};\Omega), \quad \forall\, x \in \mathbb{R}^n.\end{aligned}$$

因此得 $w \in \partial d(\bar{x};\Omega)$, 这就完成了定理的证明. □

最后给出关于可微和二次连续可微 (C^2) 函数凸性的一阶和二阶微分的刻画.

定理 2.40　设 $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ 在它的定义域 D 上是一可微函数, 这里 D 是一开凸集, 则 f 是凸的当且仅当

$$\langle \nabla f(u), x-u\rangle \leqslant f(x) - f(u), \quad \forall x, u \in D. \tag{2.21}$$

证明 必要性由命题 2.36 可得. 为证反向的结论成立, 假设式 (2.21) 成立, 并任取 $x_1, x_2 \in D, t \in (0,1)$. 记 $x_t := tx_1 + (1-t)x_2$, 由 D 的凸性得 $x_t \in D$. 则

$$\langle \nabla f(x_t), x_1 - x_t \rangle \leqslant f(x_1) - f(x_t), \quad \langle \nabla f(x_t), x_2 - x_t \rangle \leqslant f(x_2) - f(x_t).$$

进一步有

$$t\langle \nabla f(x_t), x_1 - x_t \rangle \leqslant tf(x_1) - tf(x_t)$$

和

$$(1-t)\langle \nabla f(x_t), x_2 - x_t \rangle \leqslant (1-t)f(x_2) - (1-t)f(x_t).$$

将这两个不等式相加, 得

$$0 \leqslant tf(x_1) + (1-t)f(x_2) - f(x_t),$$

这保证 $f(x_t) \leqslant tf(x_1) + (1-t)f(x_2)$, 因此证明了 f 的凸性. □

引理 2.41 设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸的且在其定义域的一个包含 $\bar{x}$ 的开子集上是二次连续可微的, 则有

$$\langle \nabla^2 f(\bar{x})u, u \rangle \geqslant 0, \quad \forall\, u \in \mathbb{R}^n,$$

其中 $\nabla^2 f(\bar{x})$ 是 f 在 $\bar{x}$ 的 Hesse 矩阵.

证明 令 $A := \nabla^2 f(\bar{x})$, 为一对称矩阵, 则

$$\lim_{h \to 0} \frac{f(\bar{x}+h) - f(\bar{x}) - \langle \nabla f(\bar{x}), h \rangle - \dfrac{1}{2}\langle Ah, h \rangle}{\|h\|^2} = 0. \tag{2.22}$$

由式 (2.22) 得对任意的 $\varepsilon > 0$, 存在 $\delta > 0$ 使得

$$-\varepsilon\|h\|^2 \leqslant f(\bar{x}+h) - f(\bar{x}) - \langle \nabla f(\bar{x}), h \rangle - \frac{1}{2}\langle Ah, h \rangle \leqslant \varepsilon\|h\|^2$$

对任意的 $\|h\| \leqslant \delta$ 成立.

根据命题 2.36, 有

$$f(\bar{x}+h) - f(\bar{x}) - \langle \nabla f(\bar{x}), h \rangle \geqslant 0.$$

合并上面不等式, 得

$$-\varepsilon\|h\|^2 \leqslant \frac{1}{2}\langle Ah, h \rangle, \quad \text{对任意的 } \|h\| \leqslant \delta \text{ 成立}. \tag{2.23}$$

进一步注意到对任意的 $0 \neq u \in \mathbb{R}^n$, 元素 $h := \delta \dfrac{u}{\|u\|}$ 满足 $\|h\| \leqslant \delta$, 代入式 (2.23) 得估计

$$-\varepsilon\delta^2 \leqslant \frac{1}{2}\delta^2 \left\langle A\frac{u}{\|u\|}, \frac{u}{\|u\|} \right\rangle.$$

因此有 (当 $u=0$ 时的情形是平凡的)

$$-2\varepsilon\|u\|^2 \leqslant \langle Au, u\rangle, \quad \forall u \in \mathbb{R}^n,$$

令 $\varepsilon \downarrow 0$, 由上式得 $0 \leqslant \langle Au, u\rangle, \forall u \in \mathbb{R}^n$. □

现在可以证明二次连续可微函数凸性的二阶刻画了.

定理 2.42 设函数 $f:\mathbb{R}^n \to \overline{\mathbb{R}}$ 为其定义域 D 上的二次连续可微函数, 其中 D 为 $\mathbb{R}^n$ 的开凸子集, 则 f 是凸的当且仅当 $\nabla^2 f(x)$ 对任意 $\bar{x} \in D$ 都是半正定的.

证明 考虑到引理 2.41, 只需证明如果对任意的 $\bar{x} \in D, \nabla^2 f(\bar{x})$ 是半正定的, 则 f 是凸的. 为此, 对任意的 $x_1, x_2 \in D$, 定义 $x_t := tx_1 + (1-t)x_2$, 考虑函数

$$\varphi(t) := f(tx_1 + (1-t)x_2) - tf(x_1) - (1-t)f(x_2), \quad t \in \mathbb{R}.$$

显然 φ 在包含 (0,1) 的开区间 I 上有定义, 则对任意的 $t \in I$, $\varphi'(t) = \langle \nabla f(x_t), x_1 - x_2\rangle - f(x_1) + f(x_2)$, 且由于 $\nabla^2 f(x_t)$ 是半正定的, 有 $\varphi''(t) = \langle \nabla^2 f(x_t)(x_1 - x_2), x_1 - x_2\rangle \geqslant 0$. 由推论 1.46 得 φ 在 I 上是凸的. 由于 $\varphi(0) = \varphi(1) = 0$, 对任意的 $t \in (0,1)$, 有

$$\varphi(t) = \varphi(t(1) + (1-t)0) \leqslant t\varphi(1) + (1-t)\varphi(0) = 0,$$

这也蕴涵着不等式

$$f(tx_1 + (1-t)x_2) \leqslant tf(x_1) + (1-t)f(x_2),$$

这就证明了函数 f 在它的定义域上是凸的, 从而在 $\mathbb{R}^n$ 上是凸的. □

2.5 基本运算法则

这里我们推导关于凸次微分式 (2.13) 及其奇异版本 (2.12) 的基本运算法则. 首先注意到关于数乘的自明法则.

命题 2.43 对凸函数 $f:\mathbb{R}^n \to \overline{\mathbb{R}}$, 其中 $\bar{x} \in \operatorname{dom} f$, 有

$$\partial(\alpha f)(\bar{x}) = \alpha\partial f(\bar{x}), \quad \partial^\infty(\alpha f)(\bar{x}) = \alpha\partial^\infty f(\bar{x}), \quad \forall\, \alpha > 0.$$

现建立关于两种结构式 (2.13) 和结构式 (2.12) 的基本的次微分和法则, 它们在凸分析中起着至关重要的作用. 我们给出关于这个结果的几何证明, 将其归结为集合法向量的交法则, 而这个结果是基于集合极点系统的凸分离定理的. 注意到我们的变分方法不需要关于所论函数的任何闭性/下半连续性假设.

定理 2.44 设 $f_i:\mathbb{R}^n \to \overline{\mathbb{R}}, i = 1, 2$ 是凸函数, 且设 $\bar{x} \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$. 假设奇异次微分规范条件

$$\partial^\infty f_1(\bar{x}) \cap [-\partial^\infty f_2(\bar{x})] = \{0\}, \tag{2.24}$$

则有次微分和法则

$$\partial(f_1 + f_2)(\bar{x}) = \partial f_1(\bar{x}) + \partial f_2(\bar{x}), \quad \partial^\infty(f_1 + f_2)(\bar{x}) = \partial^\infty f_1(\bar{x}) + \partial^\infty f_2(\bar{x}). \tag{2.25}$$

证明 我们只证明式 (2.25) 中第一个法则成立. 第二个结果的证明是类似的. 任取次梯度 $v \in \partial(f_1 + f_2)(\bar{x})$ 并定义集合

$$\Omega_1 := \{(x, \lambda_1, \lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid \lambda_1 \geqslant f_1(x)\},$$

$$\Omega_2 := \{(x, \lambda_1, \lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid \lambda_2 \geqslant f_2(x)\}.$$

首先证明包含关系

$$(v, -1, -1) \in N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_1 \cap \Omega_2) \tag{2.26}$$

成立. 对任意的 $(x, \lambda_1, \lambda_2) \in \Omega_1 \cap \Omega_2$, 有 $\lambda_1 \geqslant f_1(x)$ 和 $\lambda_2 \geqslant f_2(x)$, 则

$$\begin{aligned} &\langle v, x - \bar{x}\rangle + (-1)(\lambda_1 - f_1(\bar{x})) + (-1)(\lambda_2 - f_2(\bar{x})) \\ \leqslant &\langle v, x - \bar{x}\rangle - (f_1(x) + f_2(x) - f_1(\bar{x}) - f_2(\bar{x})) \leqslant 0, \end{aligned}$$

其中, 最后的估计成立归因于 $v \in \partial(f_1 + f_2)(\bar{x})$. 因此式 (2.26) 成立.

为在式 (2.26) 中应用交法则, 首先验证所假设的规范条件式 (2.24) 蕴涵

$$N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_1) \cap [-N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_2)] = \{0\}, \tag{2.27}$$

这是为了在式 (2.26) 情形下应用推论 2.16 所需的基本规范条件 (2.9). 事实上, 由 $\Omega_1 = \operatorname{epi} f_1 \times \mathbb{R}$ 可得

$$N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_1) = N((\bar{x}, f_1(\bar{x})); \operatorname{epi} f_1) \times \{0\},$$

并且注意到类似的表示

$$N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_2) = \{(v, 0, \gamma) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid (v, \gamma) \in N((\bar{x}, f_2(\bar{x})); \operatorname{epi} f_2)\}.$$

进一步, 取定元素

$$(v, \gamma_1, \gamma_2) \in N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_1) \cap [-N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_2)],$$

则 $\gamma_1 = \gamma_2 = 0$, 从而

$$(v, 0) \in N((\bar{x}, f_1(\bar{x})); \operatorname{epi} f_1), \quad (-v, 0) \in N((\bar{x}, f_2(\bar{x})); \operatorname{epi} f_2).$$

由奇异次微分的定义得

$$v \in \partial^\infty f_1(\bar{x}) \cap [-\partial^\infty f_2(\bar{x})] = \{0\}.$$

因此 $(v, \gamma_1, \gamma_2) = (0, 0, 0) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$. 在式 (2.26) 中应用推论 2.16 得

$$N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_1 \cap \Omega_2) = N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_1) + N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_2).$$

这蕴涵着

$$(v, -1, -1) = (v_1, -\gamma_1, 0) + (v_2, 0, -\gamma_2),$$

其中

$$(v_1, -\gamma_1) \in N((\bar{x}, f_1(\bar{x})); \operatorname{epi} f_1), \quad (v_2, -\gamma_2) \in N((\bar{x}, f_2(\bar{x})); \operatorname{epi} f_2),$$

于是得 $\gamma_1 = \gamma_2 = 1$ 且 $v = v_1 + v_2$, 这里 $v_i \in \partial f_i(\bar{x})$, $i = 1, 2$. 这就证明了式 (2.25) 中第一个次微分和法则中的包含关系“$\subset$”.

其中反向的包含关系由次微分的定义易得. 事实上, 任意的 $v \in \partial f_1(\bar{x}) + \partial f_2(\bar{x})$ 可以表示为 $v = v_1 + v_2$, 其中 $v_i \in \partial f_i(\bar{x})$, $i = 1, 2$. 于是, 对任意的 $x \in \mathbb{R}^n$ 有

$$\begin{aligned}\langle v, x - \bar{x} \rangle &= \langle v_1, x - \bar{x} \rangle + \langle v_2, x - \bar{x} \rangle \\ &\leqslant f_1(x) - f_1(\bar{x}) + f_2(x) - f_2(\bar{x}) = (f_1 + f_2)(x) - (f_1 + f_2)(\bar{x}),\end{aligned}$$

这蕴涵着 $v \in \partial (f_1 + f_2)(\bar{x})$, 这就完成了定理的证明. □

注意到根据命题 2.23, 规范条件式 (2.24) 可写为

$$N(\bar{x}; \operatorname{dom} f_1) \cap [-N(\bar{x}; \operatorname{dom} f_2)] = \{0\}.$$

下面给出定理 2.44 的一些有用的推论. 在下面推论中关于 (基本) 次微分 (2.13) 的和法则称为 Moreau-Rockafellar 定理.

推论 2.45 设 $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, 2$ 是凸函数且存在 $u \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$, 使得 f_1 在 u 处是连续的或 f_2 在 u 处是连续的, 则

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x) \tag{2.28}$$

对任意的 $x \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$ 成立. 进一步, 如果函数 f_i 在 $\mathbb{R}^n$ 上都是有限的, 则和法则式 (2.28) 对任意的 $x \in \mathbb{R}^n$ 成立.

证明 为确定起见, 假设 f_1 在 $u \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$ 处连续, 这蕴涵着 $u \in \operatorname{int}(\operatorname{dom} f_1) \cap \operatorname{dom} f_2$, 由推论 2.18 的证明可得

$$N(x; \operatorname{dom} f_1) \cap [-N(x; \operatorname{dom} f_2)] = \{0\}, \quad \forall x \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2.$$

因此结论由定理 2.44 直接可得. □

一个关于有限多函数的和的推论利用数学归纳法由定理 2.44 可得, 归纳法第二步中所需的规范条件成立. 其中由式 (2.25) 中奇异次微分和法则可知其中数学归纳法第二步中所需的规范条件成立. □

推论 2.46 考虑有限多个凸函数 $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}, i = 1, \cdots, m$. 下列断言成立:

(i) 设 $\bar{x} \in \bigcap_{i=1}^m \operatorname{dom} f_i$, 并设蕴涵关系

$$\left[\sum_{i=1}^m v_i = 0, v_i \in \partial^\infty f_i(\bar{x})\right] \Rightarrow [v_i = 0, i = 1, \cdots, m] \tag{2.29}$$

成立, 则有次微分和法则

$$\partial \left(\sum_{i=1}^m f_i\right)(\bar{x}) = \sum_{i=1}^m \partial f_i(\bar{x}), \quad \partial^\infty \left(\sum_{i=1}^m f_i\right)(\bar{x}) = \sum_{i=1}^m \partial f_i(\bar{x}).$$

(ii) 假设存在 $u \in \bigcap_{i=1}^m \operatorname{dom} f_i$ 使得所有 (可能除一个外) 函数 f_i 在 u 处是连续的, 则有等式

$$\partial \left(\sum_{i=1}^m f_i\right)(x) = \sum_{i=1}^m \partial f_i(x)$$

对任意定义域的公共点 $x \in \bigcap_{i=1}^m \operatorname{dom} f_i$ 成立.

下面的结果不仅保证凸函数的次可微性, 而且证明了其在定义域内点处的次微分的紧性.

命题 2.47 对任意凸函数 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, 任意的 $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$, 其次微分集合 $\partial f(\bar{x})$ 是非空的且在 $\mathbb{R}^n$ 中是紧的.

证明 首先证明次微分的有界性. 由推论 2.27 知 f 在 $\bar{x}$ 附近是具有某常数 ℓ 的局部 Lipschitz 连续的, 即存在一个正数 η 使得

$$|f(x) - f(y)| \leqslant \ell\|x - y\|, \quad \forall\, x, y \in \mathbb{B}(\bar{x}; \eta).$$

则对任意的次梯度 $v \in \partial f(\bar{x})$, 任意的元素 $h \in \mathbb{R}^n$, $\|h\| \leqslant \delta$, 有

$$\langle v, h\rangle \leqslant f(\bar{x} + h) - f(\bar{x}) \leqslant \ell\|h\|.$$

这蕴涵着 $\|v\| \leqslant \ell$. $\partial f(\bar{x})$ 的闭性由定义直接可得, 由此得到次梯度集合的紧性.

余下需证 $\partial f(\bar{x}) \neq \varnothing$. 由于 $(\bar{x}, f(\bar{x})) \in \operatorname{bd}(\operatorname{epi} f)$, 利用推论 2.14 可推得 $N((\bar{x}, f(\bar{x})); \operatorname{epi} f) \neq \{0\}$. 取这样的一个元素 $(0,0) \neq (v, -\gamma) \in N((\bar{x}, f(\bar{x})); \operatorname{epi} f)$, 并且根据定义得

$$\langle v, x - \bar{x}\rangle - \gamma(\lambda - f(\bar{x})) \leqslant 0, \quad \forall (x, \lambda) \in \operatorname{epi} f.$$

在不等式中取 $x := \bar{x}, \lambda = f(\bar{x}) + 1$ 得 $\gamma \geqslant 0$. 如果 $\gamma = 0$, 则根据定理 2.29 有 $v \in \partial^\infty f(\bar{x}) = \{0\}$, 这是一个矛盾. 因此 $\gamma > 0$ 且 $(v/\gamma, -1) \in N((\bar{x}, f(\bar{x})); \operatorname{epi} f)$. 因此 $v/\gamma \in \partial f(\bar{x})$, 这就完成了证明. □

给定 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 和任意的 $\alpha \in \mathbb{R}$, 定义水平集

$$\Omega_\alpha := \{\bar{x} \in \mathbb{R}^n \mid f(x) \leqslant \alpha\}.$$

命题 2.48　设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸的, 设对某 $\alpha \in \mathbb{R}$ $\bar{x} \in \Omega_\alpha$ 且 $f(\bar{x}) = \alpha$. 如果 $0 \notin \partial f(\bar{x})$, 则水平集 Ω_α 在 $\bar{x}$ 处的法锥可由

$$N(\bar{x}; \Omega_\alpha) = \partial^\infty f(\bar{x}) \cup \mathbb{R}_> \partial f(\bar{x}) \tag{2.30}$$

计算, 其中 $\mathbb{R}_>$ 表示所有正实数的集合.

证明　考虑集合 $\Theta := \mathbb{R}^n \times \{\alpha\} \subset \mathbb{R}^{n+1}$, 并注意到

$$\Omega_\alpha \times \{\alpha\} = \operatorname{epi} f \cap \Theta.$$

下面验证关系

$$N((\bar{x}, \alpha); \operatorname{epi} f) \cap [-N((\bar{x}, \alpha); \Theta)] = \{0\}, \tag{2.31}$$

即对集合 $\operatorname{epi} f$ 和 Θ 规范条件式 (2.9) 成立. 事实上, 从式 (2.31) 左边集合中取 $(v, -\lambda)$, 由 Θ 的结构可推得 $v = 0$ 和 $\lambda \geqslant 0$. 若 $\lambda > 0$, 则

$$\frac{1}{\lambda}(v, -\lambda) = (0, -1) \in N((\bar{x}, \alpha); \operatorname{epi} f),$$

因此 $0 \in \partial f(\bar{x})$, 这与该命题的假设矛盾. 故有 $(v, -\lambda) = (0, 0)$, 利用推论 2.16 的交法则可得

$$N(\bar{x}; \Omega_\alpha) \times \mathbb{R} = N((\bar{x}, \alpha); \Omega_\alpha \times \{\alpha\}) = N((\bar{x}, \alpha); \operatorname{epi} f) + (\{0\} \times \mathbb{R}).$$

现取定 $v \in N(\bar{x}; \Omega_\alpha)$, 则 $(v, 0) \in N(\bar{x}; \Omega_\alpha) \times \mathbb{R}$. 这蕴涵着存在 $\lambda \geqslant 0$ 使得 $(v, -\lambda) \in N((\bar{x}, \alpha); \operatorname{epi} f)$ 和 $(v, 0) = (v, -\lambda) + (0, \lambda)$. 如果 $\lambda = 0$, 则 $v \in \partial^\infty f(\bar{x})$. 当 $\lambda > 0$ 时, 则得 $v \in \lambda \partial f(\bar{x})$, 因此证明了式 (2.30) 中的包含关系"$\subset$".

最后证明式 (2.30) 中反向的包含关系. 由于 $\Omega_\alpha \subset \operatorname{dom} f$, 有

$$\partial^\infty f(\bar{x}) = N(\bar{x}; \operatorname{dom} f) \subset N(\bar{x}; \Omega_\alpha).$$

任取 $v \in \mathbb{R}_> \partial f(\bar{x})$, 找到 $\lambda > 0$ 使得 $v \in \lambda \partial f(\bar{x})$. 任取 $x \in \Omega_\alpha$, 则有 $f(x) \leqslant \alpha = f(\bar{x})$. 因此有

$$\langle \lambda^{-1} v, x - \bar{x} \rangle \leqslant f(x) - f(\bar{x}) \leqslant 0,$$

这蕴涵着 $v \in N(\bar{x};\Omega_\alpha)$, 从而完成了证明. □

推论 2.49 在命题 2.48 的条件下假设 f 在 $\bar{x}$ 处是连续的, 且 $0 \notin \partial f(\bar{x})$, 则有表示

$$N(\bar{x};\Omega_\alpha) = \mathbb{R}_+\partial f(\bar{x}) = \bigcup_{\lambda \geqslant 0} \lambda\partial f(\bar{x}), \quad \alpha \in \mathbb{R}.$$

证明 由于 f 在 $\bar{x}$ 是连续的, 由定理 2.29 可得 $\partial^\infty f(\bar{x}) = \{0\}$, 则该推论的结果由命题 2.48 直接可得. □

现在我们导出一些凸函数的次微分其他的运算法则. 简洁起见, 只给出关于基本次微分式 (2.13) 的结果, 奇异次微分的相应结果留作练习, 见 2.11 节.

下面为得到链式法则, 首先建立一个几何引理.

引理 2.50 设 $B : \mathbb{R}^n \to \mathbb{R}^p$ 是仿射映射由 $B(x) = A(x) + b$ 给出的仿射映射, 其中 $A : \mathbb{R}^n \to \mathbb{R}^p$ 是线性映射, 且 $b \in \mathbb{R}^p$. 则对任意的 $(\bar{x}, \bar{y}) \in \operatorname{gph} B$, 有

$$N((\bar{x},\bar{y});\operatorname{gph} B) = \{(u,v) \in \mathbb{R}^n \times \mathbb{R}^p \mid u = -A^*v\}.$$

证明 显然集合 $\operatorname{gph} B$ 是凸的, 而且 $(u,v) \in N((\bar{x},\bar{y});\operatorname{gph} B)$ 当且仅当

$$\langle u, x - \bar{x}\rangle + \langle v, B(x) - B(\bar{x})\rangle \leqslant 0, \quad \forall x \in \mathbb{R}^n. \tag{2.32}$$

由定义直接可得

$$\begin{aligned}\langle u, x - \bar{x}\rangle + \langle v, B(x) - B(\bar{x})\rangle &= \langle u, x - \bar{x}\rangle + \langle v, A(x) - A(\bar{x})\rangle \\ &= \langle u, x - \bar{x}\rangle + \langle A^*v, x - \bar{x}\rangle \\ &= \langle u + A^*v, x - \bar{x}\rangle,\end{aligned}$$

这蕴涵着式 (2.32) 与 $\langle u + A^*v, x - \bar{x}\rangle \leqslant 0$ 的等价性, 因此蕴涵着式 (2.32) 与 $u = -A^*v$ 的等价性. □

定理 2.51 设 $f : \mathbb{R}^p \to \overline{\mathbb{R}}$ 是凸函数, $B : \mathbb{R}^n \to \mathbb{R}^p$ 与引理 2.50 中的相同, 其中对某 $\bar{x} \in \mathbb{R}^p$, $B(\bar{x}) \in \operatorname{dom} f$. 记 $\bar{y} := B(\bar{x})$, 假设

$$\ker A^* \cap \partial^\infty f(\bar{y}) = \{0\}, \tag{2.33}$$

则有次微分链式法则

$$\partial(f \circ B)(\bar{x}) = A^*(\partial f(\bar{y})) = \{A^*v \mid v \in \partial f(\bar{y})\}. \tag{2.34}$$

证明 取 $v \in \partial(f \circ B)(\bar{x})$, 构造 $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}$ 的子集

$$\Omega_1 := \operatorname{gph} B \times \mathbb{R}, \quad \Omega_2 := \mathbb{R}^n \times \operatorname{epi} f.$$

由次微分和法锥的定义得 $(v,0,-1)\in N((\bar{x},\bar{y},\bar{z});\Omega_1\cap\Omega_2)$, 其中 $\bar{z}:=f(\bar{y})$. 事实上, 任取 $(x,y,\lambda)\in\Omega_1\cap\Omega_2$, 则 $y=B(x),\lambda\geqslant f(y)$, 因此 $\lambda\geqslant f(B(x))$. 从而

$$\langle v,x-\bar{x}\rangle+0(y-\bar{y})+(-1)(\lambda-\bar{z})\leqslant\langle v,x-\bar{x}\rangle-[f(B(x))-f(B(\bar{x}))]\leqslant 0.$$

利用推论 2.19 中的交法则, 在条件 (2.33) 下有

$$(v,0,-1)\in N((\bar{x},\bar{y},\bar{z});\Omega_1)+N((\bar{x},\bar{y},\bar{z});\Omega_2),$$

于是, $(v,0,-1)=(v,-w,0)+(0,w,-1)$, 其中 $(v,-\omega)\in N((\bar{x},\bar{y});\mathrm{gph}\,B)$, $(w,-1)\in N((\bar{y},\bar{z});\mathrm{epi}\,f)$, 从而

$$v=A^*w,\quad w\in\partial f(\bar{y}),$$

这相应地蕴涵着 $v\in A^*(\partial f(\bar{y}))$, 因此证明了式 (2.34) 中的包含关系"$\subset$"成立. 反向的包含关系由次微分的定义直接可得. □

下面的推论给出了一些保证定理 2.51 中的规范式 (2.33) 的条件.

推论 2.52 在定理 2.51 的条件下假设 A 是满射或 f 在点 $\bar{y}\in\mathbb{R}^n$ 处是连续的 (当 f 是有限值的, 则后面的假设自动成立), 则有次微分链式法则 (2.34).

证明 在该推论的假设下来验证规范条件式 (2.33) 成立. 由 A 的满射性可得 $\ker A^*=\{0\}$. 在第二种情形下有 $\bar{y}\in\mathrm{int}(\mathrm{dom}\,f)$, 因此 $\partial^\infty f(\bar{y})=\{0\}$. 则规范条件式 (2.33) 成立是显然的. □

定理 2.51 的最后一个推论给出了关于集合法向量的一个链式法则.

推论 2.53 设 $\Omega\subset\mathbb{R}^p$ 是凸集, 设 $B:\mathbb{R}^n\to\mathbb{R}^p$ 是与引理 2.50 中的映射相同的仿射映射, 且满足 $\bar{y}:=B(\bar{x})\in\Omega$, 则在规范条件

$$\ker A^*\cap N(\bar{y};\Omega)=\{0\}$$

下, 有

$$N(\bar{x};B^{-1}(\Omega))=A^*(N(\bar{y};\Omega)).$$

证明 在定理 2.51 中取 $f(x)=\delta(x;\Omega)$ 则直接可得结果成立. □

给定 $f_i:\mathbb{R}^n\to\overline{\mathbb{R}},i=1,\cdots,m$, 考虑最大值函数

$$f(x):=\max_{i=1,\cdots,m}f_i(x),\quad x\in\mathbb{R}^n.\tag{2.35}$$

对 $\bar{x}\in\mathbb{R}^n$, 定义集合

$$I(\bar{x}):=\{i\in\{1,\cdots,m\}\mid f_i(\bar{x})=f(\bar{x})\}.$$

在下面的命题中假设每个函数 $f_i,i=1,\cdots,m$ 在参考点处是连续的. 更一般的版本随后给出.

命题 2.54 设 $f_i:\mathbb{R}^n\to\overline{\mathbb{R}}$, $i=1,\cdots,m$ 是凸函数, 任取点 $\bar{x}\in\bigcap_{i=1}^m \operatorname{dom} f_i$, 假设每个 f_i 在 $\bar{x}$ 处是连续的, 则有最大值法则

$$\partial(\max f_i)(\bar{x})=\operatorname{co}\bigcup_{i\in I(\bar{x})}\partial f_i(\bar{x}).$$

证明 设 $f(x)$ 是式 (2.35) 中定义的最大值函数. 显然有

$$\operatorname{epi} f=\bigcap_{i=1}^m \operatorname{epi} f_i.$$

由于对任意的 $i\notin I(\bar{x})$, $f_i(\bar{x})<f(\bar{x})=:\bar{\alpha}$, 则存在 $\bar{x}$ 的一个邻域 U 和 $\delta>0$ 使得对任意的 $(x,\alpha)\in U\times(\bar{\alpha}-\delta,\bar{\alpha}+\delta)$ 有 $f_i(x)<\alpha$. 因此 $(\bar{x},\bar{\alpha})\in\operatorname{int}(\operatorname{epi} f_i)$, 从而对这样的指标 i, 有 $N((\bar{x},\bar{\alpha});\operatorname{epi} f_i)=\{(0,0)\}$.

下证对集合 $\Omega_i:=\operatorname{epi} f_i$, $i=1,\cdots,m$, 推论 2.19 中的规范条件在 $(\bar{x},\bar{\alpha})$ 处满足. 取 $(v_i,-\lambda_i)\in N((\bar{x},\bar{\alpha});\operatorname{epi} f_i)$ 满足

$$\sum_{i=1}^m (v_i,\lambda_i)=0,$$

则对 $i\notin I(x),(v_i,-\lambda_i)=(0,0)$, 因此

$$\sum_{i\in I(\bar{x})}(v_i,-\lambda_i)=0.$$

注意到对 $i\in I(\bar{x})$ 有 $f_i(\bar{x})=f(\bar{x})=\bar{\alpha}$. 由命题 2.47 的证明知对 $i\in I(\bar{x})$, 有 $\lambda_i\geqslant 0$. 因此由等式 $\sum\limits_{i\in I(\bar{x})}\lambda_i=0$ 得 $\lambda_i=0$, 这样根据定理 2.29, 对任意的 $i\in I(\bar{x})$ 有 $v_i\in\partial^\infty f_i(x)=\{0\}$. 这就证明了 $(v_i,-\lambda_i)=(0,0)$, $i=1,\cdots,m$. 进一步, 可得

$$N((\bar{x},f(\bar{x}));\operatorname{epi} f)=\sum_{i=1}^m N((\bar{x},\bar{\alpha});\operatorname{epi} f_i)=\sum_{i\in I(\bar{x})}N((\bar{x},f_i(\bar{x}));\operatorname{epi} f_i).$$

对任意的 $v\in\partial f(\bar{x})$, 有 $(v,-1)\in N((\bar{x},f(\bar{x}));\operatorname{epi} f_i)$. 于是对 $i\in I(\bar{x})$ 可找到 $(v_i,-\lambda_i)\in N((\bar{x},f_i(\bar{x}));\operatorname{epi} f_i)$ 使得

$$(v,-1)=\sum_{i\in I(\bar{x})}(v_i,-\lambda_i).$$

因此有 $\sum\limits_{i\in I(\bar{x})}\lambda_i=1$, $v=\sum\limits_{i\in I(\bar{x})}v_i$. 如果 $\lambda_i=0$, 则 $v_i\in\partial^\infty f_i(\bar{x})=\{0\}$. 因此不失一般性, 可假设对任意的 $i\in I(\bar{x}),\lambda_i>0$. 由于 $N((\bar{x},f_i(\bar{x}));\operatorname{epi} f_i)$ 是锥, 因此有

$(v_i/\lambda_i, -1) \in N((\bar{x}, f_i(\bar{x})); \text{epi}\, f_i)$, 从而 $v_i \in \lambda_i \partial f_i(\bar{x})$. 故有表示 $v = \sum_{i \in I(\bar{x})} \lambda_i u_i$, 其中 $u_i \in \partial f_i(\bar{x})$ 且 $\sum_{i \in I(\bar{x})} \lambda_i = 1$. 因此

$$v \in \text{co} \bigcup_{i \in I(\bar{x})} \partial f_i(\bar{x}).$$

注意到反向的包含关系蕴涵于

$$\partial f_i(\bar{x}) \subset \partial f(\bar{x}), \quad \forall\, i \in I(\bar{x})$$

而这可由定义得出. 事实上, 对任意的 $i \in I(\bar{x}), v \in \partial f_i(\bar{x})$, 有

$$\langle v, x - \bar{x} \rangle \leqslant f_i(x) - f_i(\bar{x}) = f_i(x) - f(\bar{x}) \leqslant f(x) - f(\bar{x}), \quad \forall\, x \in \mathbb{R}^n,$$

这蕴涵着 $v \in \partial f(\bar{x})$, 证毕. □

对一般的情形, 定义

$$\Lambda(\bar{x}) := \left\{ (\lambda_1, \cdots, \lambda_m) \,\middle|\, \lambda_i \geqslant 0, \sum_{i=1}^{m} \lambda_i = 1, \lambda_i (f_i(\bar{x}) - f(\bar{x})) = 0 \right\}.$$

并注意到由 $\Lambda(\bar{x})$ 的定义可得

$$\Lambda(\bar{x}) = \left\{ (\lambda_1, \cdots, \lambda_m) \,\middle|\, \lambda_i \geqslant 0, \sum_{i \in I(\bar{x})} \lambda_i = 1, \;\; \text{对} \;\; i \notin I(\bar{x}) \lambda_i = 0 \right\}.$$

定理 2.55 设 $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸函数, $i = 1, \cdots, m$. 任取点 $\bar{x} \in \bigcap_{i=1}^{m} \text{dom}\, f_i$, 并假设对任意的 $i \notin I(\bar{x})$, 每个 f_i 在 $\bar{x}$ 处是连续的, 而且规范条件

$$\left[\sum_{i \in I(\bar{x})} v_i = 0, v_i \in \partial^{\infty} f_i(\bar{x}) \right] \Rightarrow [v_i = 0, \forall\, i \in I(\bar{x})]$$

成立, 则有最大值法则

$$\partial(\max f_i)(\bar{x}) = \bigcup \left\{ \sum_{i \in I(\bar{x})} \lambda_i \circ \partial f_i(\bar{x}) \,\middle|\, (\lambda_1, \cdots, \lambda_m) \in \Lambda(\bar{x}) \right\},$$

这里, 如果 $\lambda_i > 0$, 则 $\lambda_i \circ \partial f_i(\bar{x}) := \lambda_i \partial f_i(\bar{x})$, 如果 $\lambda_i = 0$, 则 $\lambda_i \circ \partial f_i(\bar{x}) := \partial^{\infty} f_i(\bar{x})$.

证明 设 $f(x)$ 是式 (2.35) 中的最大值函数. 重复命题 2.54 第一部分的证明可知, 对 $i \notin I(\bar{x}), N((\bar{x}, f(\bar{x})); \text{epi}\, f_i) = \{(0,0)\}$. 所设的规范条件允许我们利用推论 2.19 中的交法则来得到

$$N((\bar{x}, f(\bar{x})); \text{epi}\, f) = \sum_{i=1}^{m} N((\bar{x}, f(\bar{x})); \text{epi}\, f_i) = \sum_{i \in I(\bar{x})} N((\bar{x}, f_i(\bar{x})); \text{epi}\, f_i). \tag{2.36}$$

对任意的 $v \in \partial f(\bar{x})$, 有 $(v,-1) \in N((\bar{x}, f(\bar{x})); \operatorname{epi} f)$, 因此利用式 (2.36), 对 $i \in I(\bar{x})$ 可找到这样的偶对 $(v_i, -\lambda_i) \in N((\bar{x}, f_i(\bar{x})); \operatorname{epi} f)$ 使得

$$(v,-1) = \sum_{i \in I(\bar{x})} (v_i, -\lambda_i).$$

这表明 $\sum_{i \in I(\bar{x})} \lambda_i = 1, v = \sum_{i \in I(\bar{x})} v_i$ 且对任意的 $i \in I(\bar{x}), v_i \in \lambda_i \circ \partial f_i(\bar{x}), \lambda_i \geqslant 0$. 因此得

$$v \in \bigcup \left\{ \sum_{i \in I(\bar{x})} \lambda_i \circ \partial f_i(\bar{x}) \middle| (\lambda_1, \cdots, \lambda_m) \in \Lambda(\bar{x}) \right\}.$$

这就证明了最大值法则中的包含关系“$\subset$”. 反向包含关系利用表达式 (2.36) 易证. □

2.6 最优值函数的次梯度

本节主要研究式 (1.4) 中定义的一类最优值/边际函数. 为此, 首先引入和检验集值映射上导数概念, 这个概念来自变分分析, 它在集值映射的许多理论及应用方面, 特别地, 与下面要考虑的最优值函数的次微分有关的理论及应用方面起着至关重要的作用.

定义 2.56 设 $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ 是具有凸图的集值映射, 设 $(\bar{x}, \bar{y}) \in \operatorname{gph} F$. F 在 $(\bar{x}, \bar{y})$ 的上导数定义为集值映射 $D^*F(\bar{x}, \bar{y}): \mathbb{R}^p \rightrightarrows \mathbb{R}^n$

$$D^*F(\bar{x}, \bar{y})(v) := \{u \in \mathbb{R}^n \mid (u, -v) \in N((\bar{x}, \bar{y}); \operatorname{gph} F)\}, \quad v \in \mathbb{R}^p. \tag{2.37}$$

由式 (2.37) 和凸集的法锥定义式 (2.4) 知凸图映射的上导数可由下式计算:

$$D^*F(\bar{x}, \bar{y})(v) = \{u \in \mathbb{R}^n \mid \langle u, \bar{x} \rangle - \langle v, \bar{y} \rangle = \operatorname*{Max}_{(x,y) \in \operatorname{gph} F} [\langle u, x \rangle - \langle v, y \rangle]\}.$$

下面给出某些特殊情形下上导数的显式计算公式.

命题 2.57 设 $F: \mathbb{R}^n \to \mathbb{R}^p$ 是由 $F(x) := Ax + b$ 给出的单值映射, 其中 $A: \mathbb{R}^n \to \mathbb{R}^p$ 是线性映射, $b \in \mathbb{R}^p$. 对任意满足 $\bar{y} = A\bar{x} + b$ 的 $(\bar{x}, \bar{y})$, 有表示

$$D^*F(\bar{x}, \bar{y})(v) = \{A^* v\}, \quad \forall v \in \mathbb{R}^p.$$

证明 由引理 2.50 得

$$N((\bar{x}, \bar{y}); \operatorname{gph} F) = \{(u, v) \in \mathbb{R}^n \times \mathbb{R}^p \mid u = -A^* v\}.$$

因此 $u \in D^*F(\bar{x}, \bar{y})(v)$ 当且仅当 $u = A^* v$. □

命题 2.58 给定凸函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$, $\bar{x}\in\operatorname{dom}f$, 定义集值映射 $F:\mathbb{R}^n\rightrightarrows\mathbb{R}$ 为

$$F(x):=[f(x),\infty].$$

记 $\bar{y}:=f(\bar{x})$, 则有

$$D^*F(\bar{x},\bar{y})(\lambda)=\begin{cases}\lambda\partial f(\bar{x}), & \lambda>0,\\ \partial^\infty f(\bar{x}), & \lambda=0,\\ \varnothing, & \lambda<0.\end{cases}$$

特别地, 如果 f 是有限的, 则 $D^*F(\bar{x},\bar{y})(\lambda)=\lambda\partial f(\bar{x})$, $\forall\,\lambda\geqslant 0$.

证明 由 F 的定义得

$$\operatorname{gph}F=\{(x,\lambda)\in\mathbb{R}^n\times\mathbb{R}\mid\lambda\in F(x)\}=\{(x,\lambda)\in\mathbb{R}^n\times\mathbb{R}\mid\lambda\geqslant f(x)\}=\operatorname{epi}f.$$

因此 $v\in D^*F(\bar{x},\bar{y})(\lambda)$ 等价于 $(v,-\lambda)\in N((\bar{x},\bar{y});\operatorname{epi}f)$. 如果 $\lambda>0$, 则 $(v/\lambda,-1)\in N((\bar{x},\bar{y});\operatorname{epi}f)$, 这蕴涵着 $v/\lambda\in\partial f(\bar{x})$, 因此 $v\in\lambda\partial f(\bar{x})$. 当 $\lambda=0$ 时的公式由奇异次微分的定义可得. 如果 $\lambda<0$, 则 $D^*F(\bar{x},\bar{y})(\lambda)=\varnothing$, 这是因为, 根据命题 2.47 的证明, $(v,-\lambda)\in N((\bar{x},\bar{y});\operatorname{epi}f)$ 蕴涵 $\lambda\geqslant 0$. 为证命题的最后一部分成立, 注意到此时根据命题 2.23 得 $D^*F(\bar{x},\bar{y})(0)=\partial^\infty f(\bar{x})=\{0\}$, 证毕. □

下面的简单例子是非常自然的且在下面是有用的.

例 2.59 定义 $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^p$ 为 $F(x)=K$, $\forall\,x\in\mathbb{R}^n$, 这里 $K\subset\mathbb{R}^p$ 是非空凸集, 则由 $\operatorname{gph}F=\mathbb{R}^n\times K$ 和命题 2.11(i) 知, 对任意的 $(\bar{x},\bar{y})\in\mathbb{R}^n\times K$, $N((\bar{x},\bar{y});\operatorname{gph}F)=\{0\}\times N(\bar{y};K)$. 因此

$$D^*F(\bar{x},\bar{y})(v)=\begin{cases}\{0\}, & -v\in N(\bar{y};K),\\ \varnothing, & \text{其他}.\end{cases}$$

特别地, 当 $K=\mathbb{R}^p$ 时有

$$D^*F(\bar{x},\bar{y})(v)=\begin{cases}\{0\}, & v=0,\\ \varnothing, & \text{其他}.\end{cases}$$

现在利用集值映射 F 的上导数和函数 φ 的次微分, 导出一个计算在形式上由

$$\mu(x):=\inf\{\varphi(x,y)|y\in F(x)\}$$

给出的最优值/边际函数 (1.4) 的次微分公式. 首先建立下面有用的估计.

引理 2.60 设 $\mu(\cdot)$ 是由凸图映射 $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^p$ 生成的最优值函数 (1.4), 凸函数 $\varphi:\mathbb{R}^n\times\mathbb{R}^p\to\overline{\mathbb{R}}$. 假设 $\mu(x)>-\infty$, $\forall\,x\in\mathbb{R}^n$, 取定某 $\bar{x}\in\operatorname{dom}\mu$, 考虑解集

$$S(\bar{x}):=\{\bar{y}\in F(\bar{x})|\mu(\bar{x})=\varphi(\bar{x},\bar{y})\}.$$

如果 $S(\bar{x}) \neq \varnothing$, 则对任意的 $\bar{y} \in S(\bar{x})$, 有

$$\bigcup_{(u,v)\in \partial\varphi(\bar{x},\bar{y})} [u + D^*F(\bar{x},\bar{y})(v)] \subset \partial\mu(\bar{x}). \tag{2.38}$$

证明 在式 (2.38) 的左边集合中取 w, 并找到 $(u,v) \in \partial\varphi(\bar{x},\bar{y})$ 满足

$$w - u \in D^*F(\bar{x},\bar{y})(v),$$

则 $(w-u,-v) \in N((\bar{x},\bar{y});\mathrm{gph}\,F)$, 从而

$$\langle w-u, x-\bar{x}\rangle - \langle v, y-\bar{y}\rangle \leqslant 0, \quad \forall\,(x,y) \in \mathrm{gph}\,F,$$

这表明对任意的 $y \in F(x)$, 有

$$\langle w, x-\bar{x}\rangle \leqslant \langle u, x-\bar{x}\rangle + \langle v, y-\bar{y}\rangle \leqslant \varphi(x,y) - \varphi(\bar{x},\bar{y}) = \varphi(x,y) - \mu(\bar{x}).$$

这蕴涵估计

$$\langle w, x-\bar{x}\rangle \leqslant \inf_{y\in F(x)} \varphi(x,y) - \mu(\bar{x}) = \mu(x) - \mu(\bar{x}),$$

这就证明了 $w \in \partial\mu(\bar{x})$, 从而完成了式 (2.38) 的证明. □

定理 2.61 在引理 2.60 的假设下考虑最优值函数 (1.4). 对任意的 $\bar{y} \in S(\bar{x})$, 有等式

$$\partial\mu(\bar{x}) = \bigcup_{(u,v)\in \partial\varphi(\bar{x},\bar{y})} [u + D^*F(\bar{x},\bar{y})(v)], \tag{2.39}$$

其中假设规范条件

$$\partial^\infty\varphi(\bar{x},\bar{y}) \cap [-N((\bar{x},\bar{y});\mathrm{gph}\,F)] = \{(0,0)\} \tag{2.40}$$

成立, 特别地, 当 φ 在 $(\bar{x},\bar{y})$ 是连续的, 该规范条件成立.

证明 考虑到引理 2.60, 只需证明式 (2.39) 中的包含关系“$\subset$”. 为此, 取 $w \in \partial\mu(\bar{x})$, $\bar{y} \in S(\bar{x})$. 对任意的 $x \in \mathbb{R}^n$, 有

$$\begin{aligned}\langle w, x-\bar{x}\rangle &\leqslant \mu(x) - \mu(\bar{x}) = \mu(x) - \varphi(\bar{x},\bar{y})\\ &\leqslant \varphi(x,y) - \varphi(\bar{x},\bar{y}), \quad \forall y \in F(x).\end{aligned}$$

对任意的 $(x,y) \in \mathbb{R}^n \times \mathbb{R}^p$, 由上式可推得下面的不等式:

$$\langle w, x-\bar{x}\rangle + \langle 0, y-\bar{y}\rangle \leqslant \varphi(x,y) + \delta((x,y);\mathrm{gph}\,F) - [\varphi(\bar{x},\bar{y}) + \delta((\bar{x},\bar{y});\mathrm{gph}\,F)].$$

进一步, 记 $f(x,y) := \varphi(x,y) + \delta((x,y);\mathrm{gph}\,F)$, 在规范条件 (2.40) 下由定理 2.44 中的次微分和法则可推得包含关系

$$(w,0) \in \partial f(\bar{x},\bar{y}) = \partial\varphi(\bar{x},\bar{y}) + N((\bar{x},\bar{y});\mathrm{gph}\,F)$$

成立. 这表明 $(w,0) = (u_1,v_1)+(u_2,v_2)$, 其中 $(u_1,v_1) \in \partial\varphi(\bar{x},\bar{y}), (u_2,v_2) \in N((\bar{x},\bar{y});\mathrm{gph}\,F)$. 因此 $v_2 = -v_1$, 从而 $(u_2,-v_1) \in N((\bar{x},\bar{y});\mathrm{gph}\,F)$. 所以 $u_2 \in D^*F(\bar{x},\bar{y})(v_1)$, 从而

$$w = u_1 + u_2 \in u_1 + D^*F(\bar{x},\bar{y})(v_1),$$

这就完成了定理的证明. □

下面给出定理 2.61 的几个重要推论. 推论 2.6.2 涉及凸复合的链式法则.

推论 2.62　设 $f:\mathbb{R}^n \to \mathbb{R}$ 是凸函数, $\varphi:\mathbb{R} \to \overline{\mathbb{R}}$ 是非减凸函数. 取 $\bar{x} \in \mathbb{R}^n$, 记 $\bar{y} := f(\bar{x})$, 并假设 $\bar{y} \in \mathrm{dom}\,\varphi$, 则在假设 $\partial^\infty\varphi(\bar{y}) = \{0\}$ 或 $0 \notin \partial f(\bar{x})$ 成立的条件下

$$\partial(\varphi\circ f)(\bar{x}) = \bigcup_{\lambda\in\partial\varphi(\bar{y})} \lambda\partial f(\bar{x}).$$

证明　根据命题 1.39 复合 $\varphi\circ f$ 是一凸函数. 定义集值映射 $F(x) := [f(x),\infty]$, 并注意到由于 φ 是非减的, 所以

$$(\varphi\circ f)(x) = \inf_{y\in F(x)} \varphi(y).$$

注意到 $\partial^\infty\varphi(\bar{y}) = \{0\}$ 和 $0 \notin \partial f(\bar{x})$ 这两个假设中的任何一个都能保证规范条件式 (2.40) 成立, 由定理 2.61 得

$$\partial(\varphi\circ f)(\bar{x}) = \bigcup_{\lambda\in\partial\varphi(\bar{y})} D^*F(\bar{x},\bar{y})(\lambda).$$

再考虑到 φ 是非减的, 对任意的 $\lambda \in \partial\varphi(\bar{y})$ 有 $\lambda \geqslant 0$. 根据命题 2.58 得

$$\partial(\varphi\circ f)(\bar{x}) = \bigcup_{\lambda\in\partial\varphi(\bar{y})} D^*F(\bar{x},\bar{y})(\lambda) = \bigcup_{\lambda\in\partial\varphi(\bar{y})} \lambda\partial f(\bar{x}),$$

这就证明了该推论中的链式法则成立. □

下面的推论说明计算参数独立约束下最优值函数的次梯度.

推论 2.63　设 $\varphi:\mathbb{R}^n\times\mathbb{R}^p \to \overline{\mathbb{R}}$, 其中对任意的 $x \in \mathbb{R}^n$, 函数 $\varphi(x,\cdot)$ 在 $\mathbb{R}^p$ 上是下有界的. 定义

$$\mu(x) := \inf\{\varphi(x,y) \mid y \in \mathbb{R}^p\},\quad S(x) := \{y \in \mathbb{R}^p \mid \varphi(x,y) = \mu(x)\}.$$

则对 $\bar{x} \in \operatorname{dom}\mu$ 及任意的 $\bar{y} \in S(\bar{x})$, 有

$$\partial\mu(\bar{x}) = \{u \in \mathbb{R}^n \mid (u, 0) \in \partial\varphi(\bar{x}, \bar{y})\}.$$

证明 在定理 2.61 中取 $F(x) \equiv \mathbb{R}^p$, 由定理 2.61 和例 2.59 直接可得结论成立. 其中注意到, 由于 $N((\bar{x}, \bar{y}); \operatorname{gph} F) = \{(0, 0)\}$, 规范条件 (2.40) 满足. □

下面的推论给出了仿射约束问题的一般结果.

推论 2.64 设 $B: \mathbb{R}^p \to \mathbb{R}^n$ 通过线性映射 $A: \mathbb{R}^p \to \mathbb{R}^n$ 和元素 $b \in \mathbb{R}^n$ 由 $B(y) := Ay + b$ 给出, 设对任意的 $x \in \mathbb{R}^n, \varphi: \mathbb{R}^p \to \overline{\mathbb{R}}$ 在逆像集合 $B^{-1}(x)$ 上是下有界的凸函数. 定义

$$\mu(x) := \inf\{\varphi(y) \mid B(y) = x\}, \quad S(x) := \{y \in B^{-1}(x) \mid \varphi(y) = \mu(x)\},$$

并取 $\bar{x} \in \mathbb{R}^n$ 满足 $\mu(\bar{x}) < \infty$ 且 $S(\bar{x}) \neq \varnothing$. 对任意的 $\bar{y} \in S(\bar{x})$, 有

$$\partial\mu(\bar{x}) = (A^*)^{-1}(\partial\varphi(\bar{y})).$$

证明 该推论的条件与当 $F(x) = B^{-1}(x), \varphi(x, y) \equiv \varphi(y)$ 时定理 2.61 的条件相同. 于是有

$$N((\bar{x}, \bar{y}); \operatorname{gph} F) = \{(u, v) \in \mathbb{R}^n \times \mathbb{R}^p \mid -A^*u = v\},$$

从而蕴涵着上导数表示

$$D^*F(\bar{x}, \bar{y})(v) = \{u \in \mathbb{R}^n \mid A^*u = v\} = (A^*)^{-1}(v), \quad v \in \mathbb{R}^p.$$

易见此时规范条件 (2.40) 显然成立. 因此

$$\partial\mu(\bar{x}) = \bigcup_{v \in \partial\varphi(\bar{y})} [D^*F(\bar{x}, \bar{y})(v)] = (A^*)^{-1}(\partial\varphi(\bar{y})),$$

这就完成了证明. □

本节最后给出推论 2.64 的一个有用应用, 这给出了在凸分析中非常重要的下卷积次微分运算法则.

推论 2.65 给定凸函数 $f_1, f_2: \mathbb{R}^n \to \overline{\mathbb{R}}$, 考虑它们的下卷积

$$(f_1 \oplus f_2)(x) := \inf\{f_1(x_1) + f_2(x_2) \mid x_1 + x_2 = x, \, x \in \mathbb{R}^n\}.$$

假设对任意的 $x \in \mathbb{R}^n$, $(f_1 \oplus f_2)(x) > -\infty$. 取 $\bar{x} \in \operatorname{dom}(f_1 \oplus f_2)$ 和 $\bar{x}_1, \bar{x}_2 \in \mathbb{R}^n$ 满足 $\bar{x} = \bar{x}_1 + \bar{x}_2$ 且 $(f_1 \oplus f_2)(\bar{x}) = f_1(\bar{x}_1) + f_2(\bar{x}_2)$, 则有

$$\partial(f_1 \oplus f_2)(\bar{x}) = \partial f_1(\bar{x}_1) \cap \partial f_2(\bar{x}_2).$$

证明 定义 $\varphi:\mathbb{R}^n\times\mathbb{R}^n\to\overline{\mathbb{R}}, A:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}^n$ 为

$$\varphi(x_1,x_2):=f_1(x_1)+f_2(x_2),\quad A(x_1,x_2):=x_1+x_2.$$

易见对任意的 $v\in\mathbb{R}^n$ 有 $A^*v=(v,v)$, 对 $(\bar{x}_1,\bar{x}_2)\in\mathbb{R}^n\times\mathbb{R}^n$ 有 $\partial\varphi(\bar{x}_1,\bar{x}_2)=(\partial f_1(\bar{x}_1),\partial f_2(\bar{x}_2))$, 则 $v\in\partial(f_1\oplus f_2)(\bar{x})$ 当且仅当 $A^*v=(v,v)\in\partial\varphi(\bar{x}_1,\bar{x}_2)$, 这意味着 $v\in\partial f_1(\bar{x}_1)\cap\partial f_2(\bar{x}_2)$. 因此欲证的计算结果由推论 2.64 可得. □

2.7 支撑函数的次梯度

本节主要研究一类重要的非光滑凸函数, 即凸集的支撑函数的次梯度性质.

定义 2.66 给定 $\mathbb{R}^n$ 的一个非空 (不一定凸) 子集 Ω, 其支撑函数定义为

$$\sigma_\Omega(x):=\sup\{\langle x,\omega\rangle\mid\omega\in\Omega\}.\tag{2.41}$$

首先给出可由定义推出支撑函数的一些性质.

命题 2.67 下面的断言成立:

(i) 对 $\mathbb{R}^n$ 的任意非空子集 Ω, 支撑函数 σ_Ω 是次可加、正齐次的, 因此是凸的.

(ii) 对任意非空子集 $\Omega_1,\Omega_2\subset\mathbb{R}^n$, 任意的 $x\in\mathbb{R}^n$, 有

$$\sigma_{\Omega_1+\Omega_2}(x)=\sigma_{\Omega_1}(x)+\sigma_{\Omega_2}(x),\quad \sigma_{\Omega_1\cup\Omega_2}(x)=\max\{\sigma_{\Omega_1}(x),\sigma_{\Omega_2}(x)\}.$$

下面通过凸分离给出了支撑函数的次梯度的计算.

定理 2.68 设 Ω 是 $\mathbb{R}^n$ 的非空闭凸子集, 设

$$S(x):=\{p\in\Omega\mid\langle x,p\rangle=\sigma_\Omega(x)\}.$$

则对任意的 $\bar{x}\in\operatorname{dom}\sigma_\Omega$, 有次微分公式

$$\partial\sigma_\Omega(\bar{x})=S(\bar{x}).$$

证明 取定 $p\in\partial\sigma_\Omega(\bar{x})$, 由定义 (2.13) 得

$$\langle p,x-\bar{x}\rangle\leqslant\sigma_\Omega(x)-\sigma_\Omega(\bar{x}),\quad\forall\,x\in\mathbb{R}^n.\tag{2.42}$$

对任意的 $u\in\mathbb{R}^n$, 由式 (2.42) 得

$$\langle p,u\rangle=\langle p,\mu+\bar{x}-\bar{x}\rangle\leqslant\sigma_\Omega(u+\bar{x})-\sigma_\Omega(\bar{x})\leqslant\sigma_\Omega(u)+\sigma_\Omega(\bar{x})-\sigma_\Omega(\bar{x})=\sigma_\Omega(u).\tag{2.43}$$

假设 $p\notin\Omega$, 应用命题 2.1 的分离结果, 找到 $u\in\mathbb{R}^n$ 满足

$$\sigma_\Omega(u)=\sup\{\langle u,\omega\rangle\mid\omega\in\Omega\}<\langle p,u\rangle,$$

这与式 (2.43) 矛盾, 因此 $p \in \Omega$. 由式 (2.43) 还得 $\langle p, \bar{x} \rangle \leqslant \sigma_\Omega(\bar{x})$. 在式 (2.42) 中令 $x = 0$ 得 $\sigma_\Omega(\bar{x}) \leqslant \langle p, \bar{x} \rangle$. 因此 $\langle p, \bar{x} \rangle = \sigma_\Omega(\bar{x})$, 从而 $p \in S(\bar{x})$.

现取定 $p \in S(\bar{x})$, 由 $p \in \Omega$ 和 $\langle p, \bar{x} \rangle = \sigma_\Omega(\bar{x})$ 得

$$\langle p, x - \bar{x} \rangle = \langle p, x \rangle - \langle p, \bar{x} \rangle \leqslant \sigma_\Omega(x) - \sigma_\Omega(\bar{x}), \quad \forall x \in \mathbb{R}^n.$$

这表明 $p \in \partial \sigma_\Omega(\bar{x})$, 这就完成了证明. □

注意到对任意的非空闭凸集 $\Omega \subset \mathbb{R}^n$, 定理 2.68 立即蕴涵着

$$\partial \sigma_\Omega(0) = S(0) = \{\omega \in \Omega \mid \langle 0, \omega \rangle = \sigma_\Omega(0) = 0\} = \Omega. \tag{2.44}$$

本节最后一个结果给出了相对于集合包含的支撑函数的某些单调性性质.

命题 2.69 设 Ω_1 和 Ω_2 是 $\mathbb{R}^n$ 的非空闭凸子集, 则包含 $\Omega_1 \subset \Omega_2$ 成立当且仅当

$$\sigma_{\Omega_1}(v) \leqslant \sigma_{\Omega_2}(v), \quad \forall v \in \mathbb{R}^n.$$

证明 取命题中的凸子集 $\Omega_1 \subset \Omega_2$, 对任意的 $v \in \mathbb{R}^n$, 有

$$\sigma_{\Omega_1}(v) = \sup\{\langle v, x \rangle \mid x \in \Omega_1\} \leqslant \sup\{\langle v, x \rangle \mid x \in \Omega_2\} = \sigma_{\Omega_2}(v).$$

反之, 假设对任意的 $v \in \mathbb{R}^n, \sigma_{\Omega_1}(v) \leqslant \sigma_{\Omega_2}(v)$. 由于 $\sigma_{\Omega_1}(0) = \sigma_{\Omega_2}(0) = 0$. 由次微分的定义 (2.13) 得

$$\partial \sigma_{\Omega_1}(0) \subset \partial \sigma_{\Omega_2}(0),$$

根据公式 (2.44), 就有 $\Omega_1 \subset \Omega_2$. □

2.8 Fenchel 共轭

凸分析及其应用 (特别地, 在最优化中的应用) 的许多重要课题都基于对偶性. 下面的概念在对偶性研究中起着至关重要的作用.

定义 2.70 给定一个函数 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$(不一定是凸的), 它的 Fenchel 共轭 $f^* : \mathbb{R}^n \to [-\infty, \infty]$ 定义为

$$f^*(v) := \sup\{\langle v, x \rangle - f(x) \mid x \in \mathbb{R}^n\} = \sup\{\langle v, x \rangle - f(x) \mid x \in \operatorname{dom} f\}. \tag{2.45}$$

注意到在式 (2.45) 中允许 $f^*(v) = -\infty$, 而如果 $\operatorname{dom} f \neq \varnothing$, 则 $f^*(v) > -\infty, \forall v \in \mathbb{R}^n$. 由定义直接得对任意非空集合 $\Omega \subset \mathbb{R}^n$, 它的指示函数的共轭是 Ω 的支撑函数:

$$\delta^*_\Omega(v) = \sup\{\langle v, x \rangle \mid x \in \Omega\} = \sigma_\Omega(v), \quad v \in \Omega. \tag{2.46}$$

易证下面的两个命题.

命题 2.71　设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是任一函数, 不一定是凸的, $\operatorname{dom} f\neq\varnothing$, 则它的 Fenchel 共轭 f^* 在 $\mathbb{R}^n$ 是凸的.

证明　作为一族仿射函数的上确界, 函数 (2.45) 是凸的. □

命题 2.72　设 $f,g:\mathbb{R}^n\to\overline{\mathbb{R}}$ 满足对任意的 $x\in\mathbb{R}^n, f(x)\leqslant g(x)$, 则有 $f^*(v)\geqslant g^*(v), \forall v\in\mathbb{R}^n$.

证明　对任意取定的 $v\in\mathbb{R}^n$, 由式 (2.45) 得

$$\langle v,x\rangle - f(x)\geqslant \langle v,x\rangle - g(x),\quad x\in\mathbb{R}^n.$$

这蕴涵着关系

$$f^*(v)=\sup\{\langle v,x\rangle - f(x)|x\in\mathbb{R}^n\}\geqslant \sup\{\langle v,x\rangle - g(x)|x\in\mathbb{R}^n\}=g^*(v),$$

对任意的 $v\in\mathbb{R}^n$ 成立, 因此在 $\mathbb{R}^n$ 上有 $f^*\geqslant g^*$. □

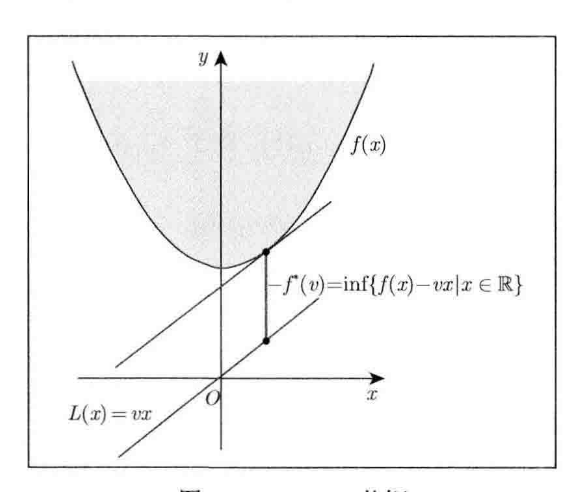


图 2.8　Fenchel 共轭

下面两个例子解释说明共轭函数计算.

例 2.73　(i) 给定 $a\in\mathbb{R}^n, b\in\mathbb{R}$, 考虑仿射函数

$$f(x):=\langle a,x\rangle + b,\quad x\in\mathbb{R}^n,$$

则由定义直接可以看到

$$f^*(v)=\begin{cases}-b, & v=a,\\ \infty, & \text{其他}.\end{cases}$$

(ii) 任给 $p > 1$, 考虑幂函数

$$f(x):=\begin{cases}\dfrac{x^p}{p}, & x \geqslant 0,\\ \infty, & \text{其他}.\end{cases}$$

共轭函数为

$$f^*(v)=\sup\left\{vx-\frac{x^p}{p}\middle| x\geqslant 0\right\}=-\inf\left\{\frac{x^p}{p}-vx\middle| x\geqslant 0\right\},\quad \forall v\in\mathbb{R}.$$

显然 $f^*(v)=0$, 如果 $v\leqslant 0$, 因为此时当 $x\geqslant 0$ 时 $vx-p^{-1}x^p\leqslant 0$. 考虑 $v>0$ 的情形, 注意到函数 $\psi_v(x):=p^{-1}x^p-vx$ 在 $(0,\infty)$ 上是凸的、可微的且 $\psi_v'(x)=x^{p-1}-v$. 因此 $\psi_v'(x)=0$ 当且仅当 $x=v^{1/(p-1)}$, 故 ψ_v 在 $x=v^{1/(p-1)}$ 处达到它的极小值. 所以共轭函数可计算如下

$$f^*(v)=\left(1-\frac{1}{p}\right)v^{p/(p-1)},\quad v\in\mathbb{R}^n.$$

取 q 满足 $q^{-1}=1-p^{-1}$, 则共轭函数可表示为

$$f^*(x)=\begin{cases}0, & v\leqslant 0,\\ \dfrac{v^q}{q}, & \text{其他}.\end{cases}$$

值得注意的是由例 2.73(ii) 的计算得

$$vx\leqslant\frac{x^p}{p}+\frac{v^q}{q},\quad \forall x,v\geqslant 0.$$

下面命题的第一个断言表明这样的关系在更一般的情形下成立. 为了表述下面第二个断言, 我们定义 f 的二次共轭为 f^* 的共轭, 即 $f^{**}(x):=(f^*)^*(x)$.

命题 2.74 设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是任一函数, $\operatorname{dom} f\neq\varnothing$, 则有

(i) $\langle v,x\rangle\leqslant f(x)+f^*(v)$, $\forall x,v\in\mathbb{R}^n$.

(ii) $f^{**}(x)\leqslant f(x)$, $\forall x\in\mathbb{R}^n$.

证明 注意到如果 $f(x)=\infty$, (i) 显然成立. 如果 $x\in\operatorname{dom} f$, 由式 (2.45) 得 $f^*(v)\geqslant\langle v,x\rangle-f(x)$, 这就证明了 (i) 成立. 它还蕴涵着

$$\sup\{\langle v,x\rangle-f^*(v)|v\in\mathbb{R}^n\}\leqslant f(x),\quad \forall x,v\in\mathbb{R}^n,$$

因此 (ii) 成立, 从而完成了证明. □

下面重要的结果揭示了凸函数的次梯度与 Fenchel 共轭之间的密切关系.

定理 2.75　对任意凸函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$, 任意的 $\bar{x}\in\operatorname{dom}f$, 有 $v\in\partial f(\bar{x})$ 当且仅当

$$f(\bar{x})+f^*(v)=\langle v,\bar{x}\rangle. \tag{2.47}$$

证明　任取 $v\in\partial f(\bar{x})$, 利用定义式 (2.13) 得

$$f(\bar{x})+\langle v,x\rangle-f(x)\leqslant\langle v,\bar{x}\rangle,\quad \forall x\in\mathbb{R}^n.$$

这蕴涵着不等式

$$f(\bar{x})+f^*(v)=f(\bar{x})+\sup\{\langle v,x\rangle-f(x)|x\in\mathbb{R}^n\}\leqslant\langle v,\bar{x}\rangle.$$

由于根据命题 2.74(i) 反向的不等式也成立, 因此式 (2.47) 成立.

反之, 假设 $f(\bar{x})+f^*(v)=\langle v,\bar{x}\rangle$. 应用命题 2.74(i), 得估计 $f^*(v)\geqslant\langle v,x\rangle-f(x)$ 对任意的 $x\in\mathbb{R}^n$ 成立, 这表明 $v\in\partial f(\bar{x})$. □

这个结果允许我们找到条件确保凸函数的二次共轭 f^{**} 等于函数本身.

命题 2.76　设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸函数, $\bar{x}\in\operatorname{dom}f$. 假设 $\partial f(\bar{x})\neq\varnothing$, 则有等式 $f^{**}(\bar{x})=f(\bar{x})$.

证明　根据命题 2.74(ii), 只需证明其中反向的不等式成立. 取 $v\in\partial f(\bar{x})$, 由定理 2.75 得 $\langle v,\bar{x}\rangle=f(\bar{x})+f^*(v)$. 这表明

$$f(\bar{x})=\langle v,\bar{x}\rangle-f^*(v)\leqslant\sup\{\langle\bar{x},v\rangle-f^*(v)|v\in\mathbb{R}^n\}=f^{**}(\bar{x}),$$

这就完成了这个命题的证明. □

考虑到命题 2.47, 则得下面的推论.

推论 2.77　设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸函数, $\bar{x}\in\operatorname{int}(\operatorname{dom}f)$, 则有等式 $f^{**}(\bar{x})=f(\bar{x})$.

本节最后证明二次共轭等式 $f=f^{**}$ 成立的一个必要且充分条件, 称为 Fenchel-Moreau 定理.

定理 2.78　设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是任一函数, $\operatorname{dom}f\neq\varnothing$, 设 $\mathcal{A}$ 是具有形式 $\varphi(x)=\langle a,x\rangle+b$, $\forall\,x\in\mathbb{R}^n$ 的所有仿射函数的集合, 其中 $a\in\mathbb{R}^n, b\in\mathbb{R}$, 记

$$\mathcal{A}(f):=\{\varphi\in\mathcal{A}|\varphi(x)\leqslant f(x),\forall\,x\in\mathbb{R}^n\},$$

则若 $\operatorname{epi}f$ 是闭凸的, 有 $\mathcal{A}(f)\neq\varnothing$, 而且, 下面论断是等价的:

(i) $\operatorname{epi}f$ 是闭凸的.

(ii) $f(x)=\sup\limits_{\varphi\in\mathcal{A}(f)}\varphi(x)$, $\forall\,x\in\mathbb{R}^n$.

(iii) $f^{**}(x)=f(x)$, $\forall\,x\in\mathbb{R}^n$.

证明 先证 $\mathcal{A}(f) \neq \varnothing$. 取定 $x_0 \in \operatorname{dom} f$, 选取 $\lambda_0 < f(x_0)$, 则 $(x_0, \lambda_0) \notin \operatorname{epi} f$. 根据命题 2.1, 存在 $(\bar{v}, \bar{\gamma}) \in \mathbb{R}^n \times \mathbb{R}$ 和 $\varepsilon > 0$ 使得

$$\langle \bar{v}, x \rangle + \bar{\gamma}\lambda < \langle \bar{v}, x_0 \rangle + \bar{\gamma}\lambda_0 - \varepsilon, \quad \forall (x, \lambda) \in \operatorname{epi} f. \tag{2.48}$$

由于对任意的 $\alpha \geqslant 0, (x_0, f(x_0) + \alpha) \in \operatorname{epi} f$, 所以有

$$\bar{\gamma}(f(\bar{x}) + \alpha) < \bar{\gamma}\lambda_0 - \varepsilon, \quad \forall \alpha \geqslant 0.$$

这意味着 $\bar{\gamma} < 0$, 若不然, 可令 $\alpha \to \infty$, 得一个矛盾. 对 $\forall x \in \operatorname{dom} f$, 因为 $(x, f(x)) \in \operatorname{epi} f$ 由式 (2.48) 得

$$\langle \bar{v}, x \rangle + \bar{\gamma} f(x) < \langle \bar{v}, x_0 \rangle + \bar{\gamma}\lambda_0 - \varepsilon, \quad \forall x \in \operatorname{dom} f.$$

所以, 如果 $x \in \operatorname{dom} f$,

$$f(x) > \left\langle \frac{\bar{v}}{\bar{\gamma}}, x_0 - x \right\rangle + \lambda_0 - \frac{\varepsilon}{\bar{\gamma}}.$$

现记 $\varphi(x) := \left\langle \dfrac{\bar{v}}{\bar{\gamma}}, x_0 - x \right\rangle + \lambda_0 - \dfrac{\varepsilon}{\bar{\gamma}}$, 则 $\varphi \in \mathcal{A}(f)$, 因此 $\mathcal{A}(f) \neq \varnothing$.

下证 (i)⇒(ii). 根据定义只需证明对任意的 $\lambda_0 < f(x_0)$, 存在 $\varphi \in \mathcal{A}(f)$, 使得 $\lambda_0 < \varphi(x_0)$. 由于 $(x_0, \lambda_0) \notin \operatorname{epi} f$, 再次利用命题 2.1 得式 (2.48). 若 $x_0 \in \operatorname{dom} f$, 上面已证 $\varphi \in \mathcal{A}(f)$. 而且, 由于 $\bar{\gamma} < 0$, 有 $\varphi(x_0) = \lambda_0 - \dfrac{\varepsilon}{\bar{\gamma}} > \lambda_0$. 现考虑 $x_0 \notin \operatorname{dom} f$ 的情形. 通过任取 $x \in \operatorname{dom} f$ 并令 $\lambda \to \infty$ 由式 (2.48) 得 $\bar{\gamma} \leqslant 0$. 如果 $\bar{\gamma} < 0$, 可以利用相同的过程得结论成立. 因此只需考虑 $\bar{\gamma} = 0$ 的情形, 此时

$$\langle \bar{v}, x - x_0 \rangle + \varepsilon < 0, \quad \forall x \in \operatorname{dom} f.$$

由于 $\mathcal{A}(f) \neq \varnothing$, 选取 $\varphi_0 \in \mathcal{A}(f)$, 并定义

$$\varphi_k(x) := \varphi_0(x) + k(\langle \bar{v}, x - x_0 \rangle + \varepsilon), \quad k \in \mathbb{N}.$$

显然, 对足够大的 k, 有 $\varphi_k \in \mathcal{A}(f), \varphi_k(x_0) = \varphi_0(x_0) + k\varepsilon > \lambda_0$. 这就证明了 (ii) 成立.

现证蕴涵关系 (ii)⇒(iii) 成立. 任取 $\varphi \in \mathcal{A}(f)$, 则 $\varphi \leqslant f$, 因此 $\varphi^{**} \leqslant f^{**}$. 应用命题 2.76, 得保证 $\varphi = \varphi^{**} \leqslant f^{**}$. 因此

$$f(x) = \sup\{\varphi(x) | \varphi \in \mathcal{A}(f)\} \leqslant f^{**}, \quad \forall x \in \mathbb{R}^n.$$

根据命题 2.74(ii) 反向的不等式 $f^{**} \leqslant f$ 也成立, 因此 $f^{**} = f$.

最后的蕴涵关系 (iii)⇒(i) 是显然的, 因为对任意函数 $g : \mathbb{R}^n \to \overline{\mathbb{R}}$, 集合 $\operatorname{epi} g^*$ 总是闭凸的. □

2.9 方向导数

接下来的课题是凸函数的方向可微性及其与次可微性的关系. 与经典分析不同, 凸分析中的方向导数结构是单侧的, 与方向的关系上不具有经典的加–减对称性.

定义 2.79 设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是增广实值函数, 设 $\bar{x}\in\mathrm{dom}\,f$, 则函数 f 在点 $\bar{x}$ 关于方向 $d\in\mathbb{R}^n$ 的方向导数是下面的极限 (如果它存在, 即为一实数, 或为 $\pm\infty$):

$$f'(\bar{x};d):=\lim_{t\to 0^+}\frac{f(\bar{x}+td)-f(\bar{x})}{t}. \tag{2.49}$$

值得一提的是, 结构 (2.49) 有时称为 f 在 $\bar{x}$ 关于方向 d 的右方向导数. 它的左方向导数定义为

$$f'_-(\bar{x};d):=\lim_{t\to 0^-}\frac{f(\bar{x}+td)-f(\bar{x})}{t}.$$

由上述定义易见

$$f'_-(\bar{x};d)=-f'(\bar{x};-d),\quad \forall\, d\in\mathbb{R}^n,$$

因此左方向导数 $f'_-(\bar{x};d)$ 的性质约化为右方向导数 (2.49) 的性质. 下面研究右方向导数.

引理 2.80 给定凸函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 且 $\bar{x}\in\mathrm{dom}\,f$, 给定 $d\in\mathbb{R}^n$, 定义

$$\varphi(t):=\frac{f(\bar{x}+td)-f(\bar{x})}{t},\quad t>0,$$

则函数 φ 在 $(0,\infty)$ 上是非减的.

证明 取定任意的数 $0<t_1<t_2$, 有表达式

$$\bar{x}+t_1d=\frac{t_1}{t_2}(\bar{x}+t_2d)+\left(1-\frac{t_1}{t_2}\right)\bar{x}.$$

由 f 的凸性得

$$f(\bar{x}+t_1d)\leqslant\frac{t_1}{t_2}f(\bar{x}+t_2d)+\left(1-\frac{t_1}{t_2}\right)f(\bar{x}),$$

于是有不等式

$$\varphi(t_1)=\frac{f(\bar{x}+t_1d)-f(\bar{x})}{t_1}\leqslant\frac{f(\bar{x}+t_2d)-f(\bar{x})}{t_2}=\varphi(t_2).$$

这就证明了 φ 在 $(0,\infty)$ 上是非减的. □

命题 2.81 对任意凸函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$, 任意 $\bar{x}\in\mathrm{dom}f$, 对任意方向 $d\in\mathbb{R}^n$, 方向导数 $f'(\bar{x};d)$ (从而它的左方向导数) 存在. 而且, 它有由引理 2.80 中定义的函数 φ 表示的表达式:

$$f'(\bar{x};d)=\inf_{t>0}\varphi(t),\quad d\in\mathbb{R}^n.$$

证明 由引理 2.80 知函数 φ 是非减的. 因此有

$$f'(\bar{x};d)=\lim_{t\to 0^+}\frac{f(\bar{x}+td)-f(\bar{x})}{t}=\lim_{t\to 0^+}\varphi(t)=\inf_{t>0}\varphi(t),$$

这就证明了该命题中所断言的结果. □

推论 2.82 如果 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸函数, 则对任意的 $\bar{x}\in \operatorname{int}(\operatorname{dom} f), d\in\mathbb{R}^n$, $f'(\bar{x};d)$ 是一实数.

证明 由定理 2.29 得 f 在 $\bar{x}$ 附近是局部 Lipschitz 连续的, 即存在 $\ell\geqslant 0$ 使得

$$\left|\frac{f(\bar{x}+td)-f(x)}{t}\right|\leqslant\frac{\ell t\|d\|}{t}=\ell\|d\|$$

对所有小的 $t>0$ 成立, 这就证明了 $|f'(\bar{x};d)|\leqslant\ell\|d\|<\infty$. □

为建立一般凸函数的方向导数和次梯度之间的关系, 需要下面有用的结果.

引理 2.83 对任意凸函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}, \bar{x}\in\operatorname{dom} f$, 有

$$f'(\bar{x};d)\leqslant f(\bar{x}+d)-f(\bar{x}),\quad \forall\, d\in\mathbb{R}^n.$$

证明 利用引理 2.80, 对其中的函数 φ, 有

$$\varphi(t)\leqslant\varphi(1)=f(\bar{x}+d)-f(\bar{x}),\quad \forall\, t\in(0,1),$$

由于 $f'(\bar{x};d)=\inf_{t>0}\varphi(t)\leqslant\varphi(1)$, 这就证明了所断言的性质. □

定理 2.84 设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸的, $\bar{x}\in\operatorname{dom} f$, 则下列是等价的:

(i) $v\in\partial f(\bar{x})$.

(ii) $\langle v,d\rangle\leqslant f'(\bar{x};d), \forall\, d\in\mathbb{R}^n$.

(iii) $f'_-(\bar{x};d)\leqslant\langle v,d\rangle\leqslant f'(\bar{x};d), \forall\, d\in\mathbb{R}^n$.

证明 任取 $v\in\partial f(\bar{x}), t>0$, 有

$$\langle v,td\rangle\leqslant f(\bar{x}+td)-f(\bar{x}),\quad \forall\, d\in\mathbb{R}^n,$$

通过两边除以 t 并取当 $t\to 0^+$ 时的极限就证明了蕴涵关系式 (i)⇒(ii) 成立. 现假设断言 (ii) 成立, 由引理 2.83 得

$$\langle v,d\rangle\leqslant f'(\bar{x};d)\leqslant f(\bar{x}+d)-f(\bar{x}),\quad \forall\, d\in\mathbb{R}^n.$$

由定义 (2.13) 它保证 $v\in\partial f(\bar{x})$, 因此断言 (i) 与 (ii) 是等价的.

显然由 (iii) 可得 (ii) 成立. 反之, 如果 (ii) 满足, 则对 $d\in\mathbb{R}^n$ 有 $\langle v,-d\rangle\leqslant f'(\bar{x};-d)$, 从而

$$f'_-(\bar{x};d)=-f'(\bar{x};-d)\leqslant\langle v,d\rangle,\quad \forall\, d\in\mathbb{R}^n.$$

这就证明 (iii) 成立, 完成了定理的证明. □

下面列出式 (2.49) 作为方向函数的一些性质.

命题 2.85　对任意凸函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}, \bar{x}\in\operatorname{dom} f$, 定义方向函数 $\psi(d):=f'(\bar{x};d)$, 它满足下面的性质:

(i) $\psi(0)=0$.

(ii) $\psi(d_1+d_2)\leqslant\psi(d_1)+\psi(d_2), \forall\, d_1,d_2\in\mathbb{R}^n$.

(iii) $\psi(\alpha d)=\alpha\psi(d), \forall\, d\in\mathbb{R}^n, \alpha>0$.

(iv) 进一步, 如果 $\bar{x}\in\operatorname{int}(\operatorname{dom} f)$, 则 ψ 在 $\mathbb{R}^n$ 上是有限的.

证明　由定义 (2.49) 直接推得性质 (i)-(iii). 例如, (ii) 成立归因于关系

$$\begin{aligned}\psi(d_1+d_2)&=\lim_{t\to0^+}\frac{f(\bar{x}+t(d_1+d_2))-f(\bar{x})}{t}\\&=\lim_{t\to0^+}\frac{f\left(\dfrac{\bar{x}+2td_1+\bar{x}+2td_2}{2}\right)-f(\bar{x})}{t}\\&\leqslant\lim_{t\to0^+}\frac{f(\bar{x}+2td_1)-f(\bar{x})}{2t}+\lim_{t\to0^+}\frac{f(\bar{x}+2td_2)-f(\bar{x})}{2t}\\&=\psi(d_1)+\psi(d_2).\end{aligned}$$

因此余下需验证对任意的 $d\in\mathbb{R}^n$, 当 $\bar{x}\in\operatorname{int}(\operatorname{dom} f)$ 时, $\psi(d)$ 是有限的. 为此, 选取 $\alpha>0$ 充分小使得 $\bar{x}+\alpha d\in\operatorname{dom} f$. 由引理 2.83 得

$$\psi(\alpha d)=f'(\bar{x};\alpha d)\leqslant f(\bar{x}+\alpha d)-f(\bar{x})<\infty.$$

利用 (iii) 得 $\psi(d)<\infty$. 进一步, 由 (i) 与 (ii) 得

$$0=\psi(0)=\psi(d+(-d))\leqslant\psi(d)+\psi(-d),\quad d\in\mathbb{R}^n,$$

这蕴涵着 $\psi(d)\geqslant-\psi(-d)$. 因此 $\psi(d)>-\infty$, 从而 (iv) 成立. □

为导出本节第二个主要结果, 需要建立另一个引理.

引理 2.86　下面有凸函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 和命题 2.85 中定义的方向函数 ψ 之间有如下关系:

(i) $\partial f(\bar{x})=\partial\psi(0)$.

(ii) $\psi^*(v)=\delta_\Omega(v), \forall\, v\in\mathbb{R}^n$, 这里 $\Omega:=\partial\psi(0)$.

证明　由定理 2.84 得 $v\in\partial f(\bar{x})$ 当且仅当

$$\langle v,d-0\rangle=\langle v,d\rangle\leqslant f'(\bar{x};d)=\psi(d)=\psi(d)-\psi(0),\quad d\in\mathbb{R}^n.$$

这等价于 $v\in\partial\psi(0)$, 因此 (i) 成立.

为证 (ii) 成立, 先证对任意的 $v \in \Omega = \partial\psi(0)$, 有 $\psi^*(v) = 0$. 事实上, 有

$$\psi^*(v) = \sup\{\langle v, d\rangle - \psi(d) |\, d \in \mathbb{R}^n\} \geqslant \langle v, 0\rangle - \psi(0) = 0.$$

现任取 $v \in \partial\psi(0)$, 则

$$\langle v, d\rangle = \langle v, d - 0\rangle \leqslant \psi(d) - \psi(0) = \psi(d), \quad d \in \mathbb{R}^n,$$

从而

$$\psi^*(v) = \sup\{\langle v, d\rangle - \psi(d) |\, d \in \mathbb{R}^n\} \leqslant 0,$$

这保证对任意的 $v \in \partial\psi(0), \psi^*(v) = 0$ 成立.

余下需证如果 $v \notin \partial\psi(0)$, 则 $\psi^*(v) = \infty$. 对这样的元素 v, 找到 $d_0 \in \mathbb{R}^n$ 满足 $\langle v, d_0\rangle > \psi(d_0)$. 由于根据命题 2.85 ψ 是正齐次的, 所以

$$\begin{aligned}\psi^*(v) &= \sup\{\langle v, d\rangle - \psi(d) |\, d \in \mathbb{R}^n\} \geqslant \sup_{t>0}(\langle v, td_0\rangle - \psi(td_0))\\ &= \sup_{t>0} t(\langle v, d_0\rangle - \psi(d_0)) = \infty,\end{aligned}$$

这就完成了引理的证明. □

现在已经准备好建立任意凸函数的方向导数与次微分之间的重要关系.

定理 2.87 给定凸函数 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, 点 $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$, 则有

$$f'(\bar{x}; d) = \max\{\langle v, d\rangle \mid v \in \partial f(\bar{x})\}, \quad \forall\, d \in \mathbb{R}^n.$$

证明 根据命题 2.85 中函数 $\psi(d) = f'(\bar{x}; d)$ 的性质, 由命题 2.76 得

$$f'(\bar{x}; d) = \psi(d) = \psi^{**}(d), \quad d \in \mathbb{R}^n.$$

注意此时 ψ 是有限凸函数. 现利用引理 2.86 得 $\psi^*(v) = \delta_\Omega(v)$, 其中 $\Omega = \partial\psi(0) = \partial f(x)$. 因此由式 (2.46) 得

$$\psi^{**}(d) = \delta^*_\Omega(d) = \sup\{\langle v, d\rangle \mid v \in \Omega\}.$$

由于根据命题 2.47 集合 $\Omega = \partial f(\bar{x})$ 是紧的, 因此完成了证明. □

2.10 上确界函数的次梯度

设 T 是 $\mathbb{R}^n$ 的非空子集, $g : T \times \mathbb{R}^n \to \mathbb{R}$. 为方便起见, 还利用记号 $g_t(x) := g(t, x)$. g_t 在 T 上的上确界函数 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 定义为

$$f(x) := \sup_{t \in T} g(t, x) = \sup_{t \in T} g_t(x), \quad x \in \mathbb{R}^n. \tag{2.50}$$

如果式 (2.50) 中的上确界是可达的 (特别地, 当指标集 T 是紧的, $g(\cdot,x)$ 是连续的, 这是成立的), 则式 (2.50) 简化为最大值函数, 当 T 是有限集时, 它可写成式 (2.35) 的形式.

本节的主要目标是当函数 g_t 为凸时计算式 (2.50) 的次微分 (2.13). 注意到此时根据命题 1.43 上确界函数 (2.50) 也是凸的. 下面我们总假设对任意的 $t \in T$ 函数 $g_t : \mathbb{R}^n \to \mathbb{R}$ 是凸的, 但不再特别指出.

对任意点 $\bar{x} \in \mathbb{R}^n$, 定义活跃指标集

$$S(\bar{x}) := \{t \in T \mid g_t(\bar{x}) = f(\bar{x})\}, \tag{2.51}$$

对 $x = \bar{x}$, 如果式 (2.50) 中的上确界不是可达的, 则该指标集可能是空的. 下面首先给出关于式 (2.50) 的一个简单下次微分估计.

命题 2.88 设对任意指标集 T 上的上确界函数 (2.50) 有定义域 $\operatorname{dom} f \neq \varnothing$, 则对任意的 $\bar{x} \in \operatorname{dom} f$, 有包含关系

$$\operatorname{clco} \bigcup_{t \in S(\bar{x})} \partial g_t(\bar{x}) \subset \partial f(\bar{x}). \tag{2.52}$$

证明 如果 $S(\bar{x}) = \varnothing$, 则包含关系式 (2.52) 显然成立. 现假设 $S(\bar{x}) \neq \varnothing$, 取定 $t \in S(\bar{x})$, $v \in \partial g_t(\bar{x})$, 则有 $g_t(\bar{x}) = f(\bar{x})$, 因此

$$\langle v, x - \bar{x} \rangle \leqslant g_t(x) - g_t(\bar{x}) = g_t(x) - f(\bar{x}) \leqslant f(x) - f(\bar{x}),$$

这表明 $v \in \partial f(\bar{x})$. 由于次梯度集合 $\partial f(\bar{x})$ 是闭凸的, 因此此时包含关系 (2.52) 式成立. □

上确界/最大值函数 (2.50) 的进一步的性质涉及下面常用的连续性的推广. 它的下半连续性版本将在第 4 章中研究和应用.

定义 2.89 设 Ω 是 $\mathbb{R}^n$ 的非空子集. 称函数 $f : \Omega \to \mathbb{R}$ 在 $\bar{x} \in \Omega$ 处是上半连续的, 如果对 Ω 中收敛于 $\bar{x}$ 的任何序列 $\{x_k\}$, 有

$$\limsup_{k \to \infty} f(x_k) \leqslant f(\bar{x}).$$

称 f 在 Ω 上是上半连续的, 如果它在该集合的任意点处是上半连续的.

命题 2.90 设 T 是 $\mathbb{R}^p$ 的非空紧子集, 对任意的 $x \in \mathbb{R}^n$, 设 $g(\cdot;x)$ 在 T 上是上半连续的, 则式 (2.50) 中的 f 是一有限凸函数.

证明 对任意的 $x \in \mathbb{R}^n$, 需验证 $f(x) < \infty$. 假设相反, 对某 $\bar{x} \in \mathbb{R}^n, f(\bar{x}) = \infty$, 则存在序列 $\{t_k\} \subset T$ 使得 $g(t_k, \bar{x}) \to \infty$. 由于 T 是紧的, 不失一般性, 假设 $t_k \to \bar{t} \in T$. 由 $g(\cdot, \bar{x})$ 在 T 上的上半连续性得

$$\infty = \limsup_{k \to \infty} g(t_k, \bar{x}) \leqslant g(\bar{t}, \bar{x}) < \infty,$$

这是一个矛盾, 从而完成了该命题的证明. □

命题 2.91 设指标集 $\varnothing \neq T \subset \mathbb{R}^p$ 是紧的, 设对任意的 $x \in \mathbb{R}^n, g(\cdot; x)$ 是上半连续的, 则对任意的 $\bar{x} \in \mathbb{R}^n$, 活跃指标集 $S(\bar{x}) \subset T$ 是非空紧的. 而且凸包

$$C := \operatorname{co} \bigcup_{t \in S(\bar{x})} \partial g_t(\bar{x})$$

在 $\mathbb{R}^n$ 中也是紧的.

证明 取定 $\bar{x} \in \mathbb{R}^n$, 对上确界函数 (2.50) 由命题 2.90 得 $f(\bar{x}) < \infty$. 考虑任意序列 $\{t_k\} \subset T$ 满足 $g(t_k, \bar{x}) \to f(\bar{x})(k \to \infty)$. 根据 T 的紧性找到 $\bar{t} \in T$ 使得沿某子序列 $t_k \to \bar{t}$. 由所假设的 g 的上半连续性得

$$f(\bar{x}) = \limsup_{k \to \infty} g(t_k, \bar{x}) \leqslant g(\bar{t}, \bar{x}) \leqslant f(\bar{x}),$$

由式 (2.51) 中 S 的构造得 $\bar{t} \in S(\bar{x})$, 从而 $S(\bar{x}) \neq \varnothing$. 由于 $g(t, \bar{x}) \leqslant f(\bar{x})$, 对任意的 $t \in T$, 得表示

$$S(\bar{x}) = \{t \in T \mid g(t, \bar{x}) \geqslant f(\bar{x})\}$$

这蕴涵着 $S(\bar{x})$ 的闭性 (也因此蕴涵着它的紧性, 由于 $S(\bar{x}) \subset T$), 这是 $g(\cdot; \bar{x})$ 上半连续的直接结果.

余下需证次梯度集

$$Q := \bigcup_{t \in S(\bar{x})} \partial g_t(\bar{x})$$

在 $\mathbb{R}^n$ 中是紧的; 这立即蕴涵着所断言的 $C = \operatorname{co} Q$ 的紧性, 因为紧集的凸包是紧的 (请读者自证, 或参见推论 3.7). 为接下来证明 Q 的紧性, 首先根据命题 2.88 有 $Q \subset \partial f(\bar{x})$. 这蕴涵着集合 Ω 在 $\mathbb{R}^n$ 中是有界的, 由于根据命题 2.47 有限值凸函数的次梯度集合是紧的. 为验证 Q 的闭性, 取收敛于某 $\bar{v}$ 的序列 $\{v_k\} \subset Q$, 对任意的 $k \in \mathbb{N}$, 找到 $t_k \in S(\bar{x})$ 使得 $v_k \in \partial g_{t_k}(\bar{x})$. 由于 $S(\bar{x})$ 是紧的, 可假设 $t_k \to \bar{t} \in S(\bar{x})$, 则 $g(t_k, \bar{x}) = g(\bar{t}, \bar{x}) = f(\bar{x})$, 并且根据定义 (2.13) 有

$$\langle v_k, x - \bar{x} \rangle \leqslant g(t_k, x) - g(t_k, \bar{x}) = g(t_k, x) - g(\bar{t}, \bar{x}), \quad \forall x \in \mathbb{R}^n.$$

从而有关系

$$\begin{aligned}\langle \bar{v}, x - \bar{x} \rangle &= \limsup_{k \to \infty} \langle v_k, x - \bar{x} \rangle \leqslant \limsup_{k \to \infty} g(t_k, x) - g(\bar{t}, \bar{x}) \\ &\leqslant g(\bar{t}, x) - g(\bar{t}, \bar{x}) = g_{\bar{t}}(x) - g_{\bar{t}}(\bar{x}),\end{aligned}$$

这就证明了 $\bar{v} \in \partial g_{\bar{t}}(\bar{x}) \subset Q$, 从而完成了证明. □

下面的结果本身具有独立的意义, 它对本节主要的次微分结果而言是至关重要的.

定理 2.92 在命题 2.91 的条件下有

$$f'(\bar{x};v) \leqslant \sigma_C(v), \quad \forall\, \bar{x} \in \mathbb{R}^n,\, v \in \mathbb{R}^n,$$

即通过其中定义的集合 C 的支撑函数来表示.

证明 由命题 2.81 和推论 2.82 得 $-\infty < f'(\bar{x};v) = \inf\limits_{\lambda>0} \varphi(\lambda) < \infty$, 其中函数 φ 是引理 2.80 中定义的, 并被证明在 $(0,\infty)$ 上是非减的. 因此存在严格递减序列 $\lambda_k \downarrow 0(k \to \infty)$ 满足

$$f'(\bar{x};v) = \lim_{k\to\infty} \frac{f(\bar{x}+\lambda_k v) - f(\bar{x})}{\lambda_k}, \quad k \in \mathbb{N}.$$

对任意的 k, 选取 $t_k \in T$ 使得 $f(\bar{x}+\lambda_k v) = g(t_k, \bar{x}+\lambda_k v)$. T 的紧性允许我们找到 $\bar{t} \in T$ 使得沿某子序列 $t_k \to \bar{t}$. 下证 $\bar{t} \in S(\bar{x})$. 由于 $g(t_k;\cdot)$ 是凸的, 且对大的 k, $\lambda_k < 1$, 有

$$g(t_k,\bar{x}+\lambda_k v) = g(t_k, \lambda_k(\bar{x}+v)+(1-\lambda_k)\bar{x}) \leqslant \lambda_k g(t_k,\bar{x}+v)+(1-\lambda_k)g(t_k,\bar{x}), \tag{2.53}$$

根据 f 的连续性 (任何有限值凸函数是连续的) 和 $g(\cdot;x)$ 的上半连续性, 式 (2.53) 蕴涵着

$$f(\bar{x}) = \lim_{k\to\infty} f(\bar{x}+\lambda_k v) = \lim_{k\to\infty} g(t_k,\bar{x}+\lambda_k v) \leqslant g(\bar{t},\bar{x}).$$

因此 $f(\bar{x}) = g(\bar{t},\bar{x})$, 从而 $\bar{t} \in S(\bar{x})$. 而且, 对任意的 $\varepsilon > 0$ 和大的 k, 根据 g 的上半连续性有

$$g(t_k,\bar{x}+v) \leqslant \gamma + \varepsilon, \quad 其中 \;\; \gamma := g(\bar{t},\bar{x}+v).$$

由式 (2.53) 得

$$g(t_k,\bar{x}) \geqslant \frac{g(t_k,\bar{x}+\lambda_k v) - \lambda_k(\gamma+\varepsilon)}{1-\lambda_k} \to f(\bar{x}),$$

从而 $\lim\limits_{k\to\infty} g(t_k,\bar{x}) = f(\bar{x})$, 这是因为

$$f(\bar{x}) \leqslant \liminf_{k\to\infty} g(t_k,\bar{x}) \leqslant \limsup_{k\to\infty} g(t_k,\bar{x}) \leqslant g(\bar{t},\bar{x}) = f(\bar{x}).$$

进一步, 任取定 $\lambda > 0$, 并考虑到对任意充分大的 k, 有 $\lambda_k < \lambda$, 利用引理 2.80 推得

$$\begin{aligned}
\frac{f(\bar{x}+\lambda_k v) - f(\bar{x})}{\lambda_k} &= \frac{g(t_k,\bar{x}+\lambda_k v) - g(\bar{t},\bar{x})}{\lambda_k} \\
&\leqslant \frac{g(t_k,\bar{x}+\lambda_k v) - g(t_k,\bar{x})}{\lambda_k} \\
&\leqslant \frac{g(t_k,\bar{x}+\lambda v) - g(t_k,\bar{x})}{\lambda},
\end{aligned}$$

从而有估计

$$\begin{aligned}\limsup_{k\to\infty}\frac{f(\bar{x}+\lambda_k v)-f(\bar{x})}{\lambda_k}&\leqslant\limsup_{k\to\infty}\frac{g(t_k,\bar{x}+\lambda v)-g(t_k,\bar{x})}{\lambda}\\&\leqslant\frac{g(\bar{t},\bar{x}+\lambda v)-g(\bar{t},\bar{x})}{\lambda}\\&=\frac{g_{\bar{t}}(\bar{x}+\lambda v)-g_{\bar{t}}(\bar{x})}{\lambda},\end{aligned}$$

这蕴涵着关系

$$f'(\bar{x};v)\leqslant\inf_{\lambda>0}\frac{g_{\bar{t}}(\bar{x}+\lambda v)-g_{\bar{t}}(\bar{x})}{\lambda}=g_{\bar{t}}'(\bar{x};v).$$

最后利用定理 2.87 和命题 2.88 中的包含关系 $\partial g_{\bar{t}}(\bar{x})\subset C$ 得

$$f'(\bar{x};v)\leqslant g_{\bar{t}}'(\bar{x};v)=\sup_{w\in\partial g_{\bar{t}}(\bar{x})}\langle w,v\rangle\leqslant\sup_{w\in C}\langle w,v\rangle=\sigma_C(v),$$

这就证明了定理的结论成立. □

现在我们已经准备好建立本节的主要结果.

定理 2.93　设 f 是式 (2.50) 中所定义的, 其中指标集 $\varnothing\neq T\subset\mathbb{R}^p$ 是紧的, 对任意的 $t\in T$, 函数 $g(t,\cdot)$ 在 $\mathbb{R}^n$ 上是凸的. 而且假设对任意的 $x\in\mathbb{R}^n$ 函数 $g(\cdot,x)$ 在 T 上是上半连续的, 则有

$$\partial f(\bar{x})=\operatorname{co}\bigcup_{t\in S(\bar{x})}\partial g_t(\bar{x}).$$

证明　任取 $w\in\partial f(\bar{x})$, 由定理 2.87 和定理 2.92 得

$$\langle w,v\rangle\leqslant f'(\bar{x},v)\leqslant\sigma_C(v),\quad\forall\, v\in\mathbb{R}^n,$$

其中 C 是命题 2.91 中所定义的. 因此对集合 $\Omega=C$, 它是 $\mathbb{R}^n$ 的非空闭凸子集, 根据公式 (2.44) 得 $w\in\partial\sigma_C(0)=C$. 反向的包含关系由命题 2.88 可得. □

2.11 练　　习

练习 2.1　(i) 设 $\Omega\subset\mathbb{R}^n$ 是非空闭凸锥且设 $\bar{x}\notin\Omega$, 证明存在非零元素 $v\in\mathbb{R}^n$ 使得对任意的 $x\in\Omega$, $\langle v,x\rangle\leqslant 0$, 而且 $\langle v,\bar{x}\rangle>0$.

(ii) 设 Ω 是 $\mathbb{R}^n$ 的子空间, 设 $\bar{x}\notin\Omega$. 证明存在非零元素 $v\in\mathbb{R}^n$ 使得对任意的 $x\in\Omega$, $\langle v,x\rangle=0$ 且 $\langle v,\bar{x}\rangle>0$.

练习 2.2　设 Ω_1 和 Ω_2 是 $\mathbb{R}^n$ 的两个非空凸子集满足条件 $\operatorname{ri}\Omega_1\cap\operatorname{ri}\Omega_2=\varnothing$. 证明存在非零元素 $v\in\mathbb{R}^n$ 在式 (2.2) 的意义下分离集合 Ω_1 与 Ω_2.

练习 2.3　$\mathbb{R}^n$ 的两个非空凸子集称为正常分离的, 如果存在非零元素 $v \in \mathbb{R}^n$ 使得

$$\sup\{\langle v, x\rangle \mid x \in \Omega_1\} \leqslant \inf\{\langle v, y\rangle \mid y \in \Omega_2\},$$

$$\inf\{\langle v, x\rangle \mid x \in \Omega_1\} < \sup\{\langle v, y\rangle \mid y \in \Omega_2\}.$$

证明 Ω_1 和 Ω_2 是正常分离的当且仅当 $\operatorname{ri}\Omega_1 \cap \operatorname{ri}\Omega_2 = \varnothing$.

练习 2.4　设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸函数且在某点 $\bar{x} \in \operatorname{dom} f$ 处达到极小值. 证明集合 $\{\Omega_1, \Omega_2\}$, 其中 $\Omega_1 := \operatorname{epi} f$, $\Omega_2 := \mathbb{R}^n \times \{f(\bar{x})\}$ 构成一个极点系统.

练习 2.5　设 $\Omega_i, i = 1, 2$ 是 $\mathbb{R}^n$ 的非空凸子集, 至少有一公共点, 且满足

$$\operatorname{int}\Omega_1 \neq \varnothing \text{ 且 } \operatorname{int}\Omega_1 \cap \Omega_2 = \varnothing.$$

证明存在不同时等于零的元素 $v_1, v_2 \in \mathbb{R}^n$ 和实数 α_1, α_2 满足

$$\langle v_i, x\rangle \leqslant \alpha_i, \quad \forall\, x \in \Omega_i,\ i = 1, 2,$$

$$v_1 + v_2 = 0 \text{ 且 } \alpha_1 + \alpha_2 = 0.$$

练习 2.6　设 $\Omega := \{x \in \mathbb{R}^n \mid x \leqslant 0\}$, 证明:

$$N(\bar{x}; \Omega) = \{v \in \mathbb{R}^n \mid v \geqslant 0, \langle v, \bar{x}\rangle = 0\}, \quad \forall\, \bar{x} \in \Omega.$$

练习 2.7　设 $f : \mathbb{R}^n \to \mathbb{R}$ 是一在 $\bar{x}$ 处是可微的凸函数. 证明:

$$N((\bar{x}, f(\bar{x})); \operatorname{epi} f) = \{\lambda(\nabla f(\bar{x}), -1) \mid \lambda \geqslant 0\}.$$

练习 2.8　用数学归纳法证明推论 2.19.

练习 2.9　计算下列凸函数在它们定义域的每一点处的次微分和奇异次微分.

(i) $f(x) = |x - 1| + |x - 2|, x \in \mathbb{R}$.

(ii) $f(x) = \mathrm{e}^{|x|}, x \in \mathbb{R}$.

(iii) $f(x) = \begin{cases} x^2 - 1, & |x| \leqslant 1, \\ \infty, & \mathbb{R} \text{ 上的其他点}. \end{cases}$

(iv) $f(x) = \begin{cases} -\sqrt{1 - x^2}, & |x| \leqslant 1, \\ \infty, & \mathbb{R}\text{上的其他点}. \end{cases}$

(v) $f(x_1, x_2) = |x_1| + |x_2|, (x_1, x_2) \in \mathbb{R}^2$.

(vi) $f(x_1, x_2) = \max\{|x_1|, |x_2|\}, (x_1, x_2) \in \mathbb{R}^2$.

(vii) $f(x) = \max\{\langle a, x\rangle - b, 0\}, x \in \mathbb{R}^n, a \in \mathbb{R}^n, b \in \mathbb{R}$.

练习 2.10　设 $f:\mathbb{R}\to\mathbb{R}$ 是非减凸函数, 设 $\bar{x}\in\mathbb{R}$. 证明对 $v\in\partial f(\bar{x})$, 有 $v\geqslant 0$.

练习 2.11　用数学归纳法证明推论 2.46.

练习 2.12　设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸函数, 给出命题 2.47 中对任意元素 $\bar{x}\in\operatorname{int}(\operatorname{dom}f),\partial f(\bar{x})\neq\varnothing$ 这部分的另一种证明.

如果 $\bar{x}\in\operatorname{ri}(\operatorname{dom}f)$, 结论还成立吗?

练习 2.13　设 Ω_1 和 Ω_2 是 $\mathbb{R}^n$ 的非空凸子集, 假设 $\operatorname{ri}\Omega_1\cap\operatorname{ri}\Omega_2\neq\varnothing$. 证明

$$N(\bar{x};\Omega_1\cap\Omega_2)=N(\bar{x};\Omega_1)+N(\bar{x};\Omega_2),\quad\forall\bar{x}\in\Omega_1\cap\Omega_2.$$

练习 2.14　设 $f_1,f_2:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸函数且满足条件

$$\operatorname{ri}(\operatorname{dom}f_1)\cap\operatorname{ri}(\operatorname{dom}f_2)\neq\varnothing.$$

证明对任意的 $\bar{x}\in\operatorname{dom}f_1\cap\operatorname{dom}f_2$, 有

$$\partial(f_1+f_2)(\bar{x})=\partial f_1(\bar{x})+\partial f_2(\bar{x}),\quad\partial^\infty(f_1+f_2)(\bar{x})=\partial^\infty f_1(\bar{x})+\partial^\infty f_2(\bar{x}).$$

练习 2.15　在定理 2.51 的条件和规范条件 (2.33) 下证明

$$N((\bar{x},\bar{y},\bar{z});\Omega_1)\cap[-N((\bar{x},\bar{y},\bar{z});\Omega_2)]=\{0\}.$$

练习 2.16　在定理 2.51 的条件下计算复合 $f\circ B$ 的奇异次微分.

练习 2.17　考虑命题 1.54 条件下的复合 $g\circ h$, 任取 $\bar{x}\in\operatorname{dom}(g\circ h)$. 证明

$$\partial(g\circ h)(\bar{x})=\left\{\sum_{i=1}^{p}\lambda_i v_i\,\middle|\,(\lambda_1,\cdots,\lambda_p)\in\partial g(\bar{y}),v_i\in\partial f_i(\bar{x}),i=1,\cdots,p\right\}$$

在规范条件

$$\left[0\in\sum_{i=1}^{p}\lambda_i\partial f_i(\bar{x}),(\lambda_1,\cdots,\lambda_p)\in\partial^\infty g(\bar{y})\right]\Rightarrow[\lambda_i=0,i=1,\cdots,p]$$

下成立.

练习 2.18　计算定理 2.61 条件下最优值函数的奇异次微分.

练习 2.19　计算下列集合的支撑函数:

(i) $\Omega_1:=\left\{x=(x_1,\cdots,x_n)\in\mathbb{R}^n\,\middle|\,\|x\|_1:=\sum_{i=1}^{n}|x_i|\leqslant 1\right\}$.

(ii) $\Omega_2:=\{x=(x_1,\cdots,x_n)\in\mathbb{R}^n\,|\,\|x\|_\infty:=\max_{i=1}^{n}|x_i|\leqslant 1\}$.

练习 2.20 找到 $\mathbb{R}$ 上下列每个函数的 Fenchel 共轭:

(i) $f(x) = \mathrm{e}^x$.

(ii) $f(x) = \begin{cases} -\ln(x), & x > 0, \\ \infty, & x \leqslant 0. \end{cases}$

(iii) $f(x) = ax^2 + bx + c$, 这里 $a > 0$.

(iv) $f(x) = \delta_{\mathbb{B}}(x)$, 这里 $\mathbb{B} = [-1, 1]$.

(v) $f(x) = |x|$.

练习 2.21 设 Ω 是 $\mathbb{R}^n$ 的非空子集. 求二次共轭函数 δ_Ω^{**}.

练习 2.22 证明 $\mathbb{R}^n$ 上的函数的下列断言是正确的:

(i) $(\lambda f)^*(v) = \lambda f^*\left(\dfrac{v}{\lambda}\right)$, 其中 $\lambda > 0$.

(ii) $(f + \lambda)^*(v) = f^*(v) - \lambda$, 其中 $\lambda \in \mathbb{R}$.

(iii) $f_u^*(v) = f^*(v) + \langle u, v\rangle$, 其中 $u \in \mathbb{R}^n, f_u(x) := f(x - u)$.

练习 2.23 设 $f : \mathbb{R}^n \to \mathbb{R}$ 是凸正齐次函数. 证明 $f^*(v) = \delta_\Omega(v), \forall\, v \in \mathbb{R}^n$, 这里 $\Omega := \partial f(0)$.

练习 2.24 设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是一函数且 $\operatorname{dom} f = \varnothing$. 证明 $f^{**}(x) = f(x), \forall\, x \in \mathbb{R}^n$.

练习 2.25 设 $f : \mathbb{R} \to \mathbb{R}$ 是凸的. 证明 $\partial f(x) = [f'_-(x), f'_+(x)]$, 这里 $f'_-(x)$ 和 $f'_+(x)$ 分别代表 f 在 x 处的左、右导数.

练习 2.26 定义函数 $f : \mathbb{R} \to \overline{\mathbb{R}}$ 为

$$f(x) := \begin{cases} -\sqrt{1 - x^2}, & |x| \leqslant 1, \\ \infty, & \text{其他}, \end{cases}$$

证明 f 是凸函数并计算关于方向 $d = 1$ 时方向导数 $f'(-1; d)$ 和 $f'(1; d)$.

练习 2.27 设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸的, 设 $\bar{x} \in \operatorname{dom} f$. 对元素 $d \in \mathbb{R}^n$, 考虑引理 2.80 中定义的函数 φ. 证明 φ 在 $(-\infty, 0)$ 上是非减的.

练习 2.28 设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸的, 设 $\bar{x} \in \operatorname{dom} f$. 证明下列结论成立:

(i) $f'_-(\bar{x}; d) \leqslant f'(\bar{x}; d), \forall d \in \mathbb{R}^n$.

(ii) $f(\bar{x}) - f(\bar{x} - d) \leqslant f'_-(\bar{x}; d) \leqslant f'(\bar{x}; d) \leqslant f(\bar{x} + d) - f(\bar{x}), d \in \mathbb{R}^n$.

练习 2.29 证明命题 2.67 的断言成立.

练习 2.30 设 $\Omega \subset \mathbb{R}^n$ 是非空紧集. 利用定理 2.93, 计算支撑函数的次微分 $\partial\sigma_\Omega(0)$.

练习 2.31 设 Ω 是 $\mathbb{R}^n$ 的非空紧子集, 定义函数

$$\mu_\Omega(x) := \sup\{\|x - \omega\| \mid \omega \in \Omega\}, \quad x \in \mathbb{R}^n.$$

证明 μ_Ω 是凸函数, 并求它在任意点 $x \in \mathbb{R}^n$ 处的次微分 $\partial \mu_\Omega(x)$.

练习 2.32　利用表示

$$\|x\| = \sup\{\langle v, x\rangle \mid \|v\| \leqslant 1\}$$

和定理 2.93, 计算范数函数 $p(x) := \|x\|$ 在任意点 $\bar{x} \in \mathbb{R}^n$ 处的次微分 $\partial p(\bar{x})$.

练习 2.33　设 $d(x;\Omega)$ 是关联于 $\mathbb{R}^n$ 中某非空闭凸子集 Ω 的距离函数 (1.5).

(i) 证明由 Ω 的支撑函数表示的表达式

$$d(x;\Omega) = \sup\{\langle x, v\rangle - \sigma_\Omega(v) \mid \|v\| \leqslant 1\}$$

成立.

(ii) 计算在任意点 $\bar{x} \in \mathbb{R}^n$ 处的次微分 $\partial d(\bar{x};\Omega)$.

第 3 章　基于凸性的有名结果

本章致力于凸分析中一些值得注意的课题, 与第 2 章的次微分运算和相关结论一同构成了这个学科的主要成果, 对许多应用 (首先对凸最优化的应用) 来说这是最重要的课题.

3.1　可微性的刻画

本节定义可微性的另一经典概念, 即 Gâteaux 导数, 并证明对凸函数来说它与 Fréchet 导数一样可通过凸分析中的次微分的单点性来刻画.

定义 3.1　设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 在 $\bar{x}$ 附近是有限的. 称 f 在 $\bar{x}$ 处是 Gâteaux 可微的, 如果存在 $v\in\mathbb{R}^n$ 使得对任意的 $d\in\mathbb{R}^n$, 有

$$\lim_{t\to 0}\frac{f(\bar{x}+td)-f(\bar{x})-t\langle v,d\rangle}{t}=0,$$

此时称 v 是 f 在 $\bar{x}$ 处的 Gâteaux 导数, 并记为 $f'_G(\bar{x})$.

由定义和等式 $f'_-(\bar{x};d)=-f'(\bar{x};-d)$ 得 $f'_G(\bar{x})=v$ 当且仅当方向导数 $f'(\bar{x};d)$ 存在且相对于 d 是线性的, 满足对任意的 d, $f'(\bar{x};d)=\langle v,d\rangle$; 参见练习 3.1 及其解答. 进一步, 如果 f 在 $\bar{x}$ 处是 Gâteaux 可微, 则对任意的 $d\in\mathbb{R}^n$, 有

$$\langle f'_G(\bar{x}),d\rangle=\lim_{t\to 0}\frac{f(\bar{x}+td)-f(\bar{x})}{t}.$$

命题 3.2　设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是任一函数 (不一定是凸的), 它在 $\bar{x}\in\operatorname{dom}f$ 附近是局部 Lipschitz 连续的, 则下列断言是等价的:

(i) f 在 $\bar{x}$ 处是 Fréchet 可微的.

(ii) f 在 $\bar{x}$ 处是 Gâteaux 可微的.

证明　设 f 在 $\bar{x}$ 处是 Fréchet 可微的, 并设 $v:=\nabla f(\bar{x})$. 对取定的 $0\neq d\in\mathbb{R}^n$, 考虑到 $\|\bar{x}+td-\bar{x}\|=|t|\cdot\|d\|\to 0(t\to 0)$, 有

$$\lim_{t\to 0}\frac{f(\bar{x}+td)-f(\bar{x})-t\langle v,d\rangle}{t}=\|d\|\lim_{t\to 0}\frac{f(\bar{x}+td)-f(\bar{x})-\langle v,td\rangle}{t\|d\|}=0,$$

这就蕴涵着 f 在 $\bar{x}$ 处的 Gâteaux 可微性且 $f'_G(\bar{x})=\nabla f(\bar{x})$. 值得注意的是, 在这个蕴涵关系中不要求 f 在 $\bar{x}$ 附近的局部 Lipschitz 连续性.

现假设 f 在 $\bar{x}$ 处是 Gâteaux 可微的且 $v := f'_G(\bar{x})$, 下证

$$\lim_{h\to 0}\frac{f(\bar{x}+h)-f(\bar{x})-\langle v,h\rangle}{\|h\|}=0.$$

若相反地假设它不成立, 则可找到 $\varepsilon_0>0$, 序列 $h_k\to 0$ 且 $h_k\neq 0(k\in\mathbb{N})$ 使得

$$\frac{|f(\bar{x}+h_k)-f(\bar{x})-\langle v,h_k\rangle|}{\|h_k\|}\geqslant\varepsilon_0. \tag{3.1}$$

记 $t_k:=\|h_k\|$, $d_k:=\dfrac{h_k}{\|h_k\|}$, 则 $t_k\to 0$. 不失一般性, 可假设 $d_k\to d$ 且 $\|d\|=1$. 因此有估计

$$\varepsilon_0\leqslant\frac{|f(\bar{x}+t_kd_k)-f(\bar{x})-\langle v,t_kd_k\rangle|}{t_k}.$$

设 $\ell\geqslant 0$ 是 f 在 $\bar{x}$ 附近的一个 Lipschitz 常数. 三角不等式保证

$$\begin{aligned}
\varepsilon_0&\leqslant\frac{|f(\bar{x}+t_kd_k)-f(\bar{x})-\langle v,t_kd_k\rangle|}{t_k}\\
&\leqslant\frac{|f(\bar{x}+t_kd_k)-f(\bar{x}+t_kd)|+|\langle v,t_kd\rangle-\langle v,t_kd_k\rangle|}{t_k}\\
&\quad+\frac{|f(\bar{x}+t_kd)-f(\bar{x})-\langle v,t_kd\rangle|}{t_k}\\
&\leqslant\ell\|d_k-d\|+\|v\|\cdot\|d_k-d\|+\frac{|f(\bar{x}+t_kd)-f(\bar{x})-\langle v,t_kd\rangle|}{t_k}\to 0(k\to\infty).
\end{aligned}$$

这是一个矛盾. 因此 f 在 $\bar{x}$ 处是 Fréchet 可微的.

现在我们已经准备好得到可微性的一个漂亮的次微分刻画.

定理 3.3 设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸函数且 $\bar{x}\in\operatorname{int}(\operatorname{dom}f)$, 则下列断言是等价的:

(i) f 在 $\bar{x}$ 处是 Fréchet 可微的.

(ii) f 在 $\bar{x}$ 处是 Gâteaux 可微的.

(iii) $\partial f(\bar{x})$ 是一单点集.

证明 由定理 2.29 知 f 在 $\bar{x}$ 附近是局部 Lipschitz 连续的, 因此根据命题 3.2, 断言 (i) 与 (ii) 是等价的. 现假设 (ii) 成立, 令 $v:=f'_G(\bar{x})$, 则对任意的 $d\in\mathbb{R}^n$, $f'_-(\bar{x};d)=f'(\bar{x};d)=\langle v,d\rangle$. 注意到, 由于 $\bar{x}\in\operatorname{int}(\operatorname{dom}f)$ 得 $\partial f(\bar{x})\neq\varnothing$. 任取 $w\in\partial f(\bar{x})$, 由定理 2.84 得

$$f'_-(\bar{x};d)\leqslant\langle w,d\rangle\leqslant f'(\bar{x};d),\quad\forall\, d\in\mathbb{R}^n,$$

因此 $\langle w,d\rangle=\langle v,d\rangle$, 从而对任意的 d, $\langle w-v,d\rangle=0$. 特别地, 取 $d=w-v$, 则得 $w=v$, 从而证明了 $\partial f(\bar{x})$ 是单点集.

最后假设 (iii) 成立且 $\partial f(\bar{x}) = \{v\}$, 下证 (ii) 成立. 根据定理 2.87, 对任意的 $d \in \mathbb{R}^n$, 有

$$f'(\bar{x}; d) = \sup_{w \in \partial f(\bar{x})} \langle w, d \rangle = \langle v, d \rangle.$$

这蕴涵着 f 在 $\bar{x}$ 处是 Gâteaux 可微的, 从而也是 Fréchet 可微的. □

3.2 Carathéodory 定理和 Farkas 引理

本节首先给出关于集合锥包和凸包的基本结果, 然后导出凸几何最著名的结果之一, 这个结果由 Constantin Carathéodory 于 1911 年建立. 回顾一个集合 Ω 称为一个锥, 如果 $\lambda x \in \Omega, \forall \lambda \geqslant 0, x \in \Omega$. 由定义得如果 Ω 是非空锥, 则 $0 \in \Omega$. 给定非空集合 $\Omega \in \mathbb{R}^n$, 称 K_Ω 是由 Ω 生成的凸锥, 或称 Ω 的凸锥包, 如果它是包含 Ω 的所有凸锥的交.

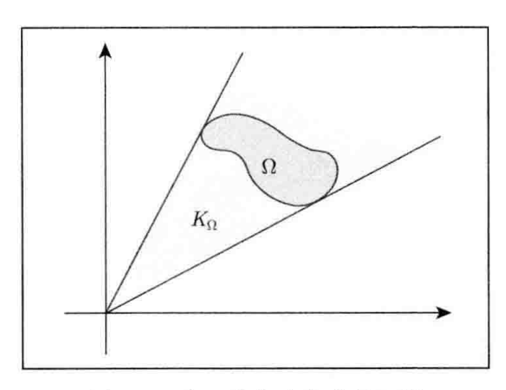


图 3.1　由一个集合生成的凸锥

命题 3.4　设 Ω 是 $\mathbb{R}^n$ 的非空子集, 则由 Ω 生成的凸锥有下面的表示:

$$K_\Omega = \left\{ \sum_{i=1}^{m} \lambda_i a_i \,\middle|\, \lambda_i \geqslant 0, a_i \in \Omega, m \in \mathbb{N} \right\}. \tag{3.2}$$

特别地, 如果集合 Ω 是凸的, 则 $K_\Omega = \mathbb{R}_+ \Omega$.

证明　记 C 是式 (3.2) 右边的集合, 注意到它是包含 Ω 的凸锥. 余下需证明对包含 Ω 的任何凸锥 K, 有 $C \subset K$. 为此，取定这样的一个锥, 并构造组合 $x = \sum_{i=1}^{m} \lambda_i a_i$, 其中 $\forall \lambda_i \geqslant 0$, $a_i \in \Omega$ 且 $m \in \mathbb{N}$. 如果对任意的 i, $\lambda_i = 0$, 则 $x = 0 \in K$. 否则, 假设对某个 i, $\lambda_i > 0$, 记 $\lambda := \sum_{i=1}^{m} \lambda_i > 0$. 因此

$$x = \lambda\left(\sum_{i=1}^{m}\frac{\lambda_i}{\lambda}a_i\right) \in K,$$

这就有 $C \subset K$, 因此 $K_\Omega = C$. 该命题最后一个断言是显然的. □

命题 3.5 对任意非空集合 $\Omega \subset \mathbb{R}^n$ 和任意的 $x \in K_\Omega \setminus \{0\}$, 有

$$x = \sum_{i=1}^{m}\lambda_i a_i, \quad 其中 \quad \lambda_i > 0, \quad a_i \in \Omega, \quad i = 1, \cdots, m \quad 且 \quad m \leqslant n.$$

证明 设 $x \in K_\Omega \setminus \{0\}$. 由命题 3.4 得 $x = \sum_{i=1}^{m}\mu_i a_i$, 其中某个 $\mu_i > 0$, $a_i \in \Omega$, $i = 1, \cdots, m$, $m \in \mathbb{N}$. 如果元素 $a_1, \cdots, a_m$ 是线性相关的, 则存在不全为 0 的数 $\gamma_i \in \mathbb{R}$, 使得

$$\sum_{i=1}^{m}\gamma_i a_i = 0.$$

可假设 $I := \{i = 1, \cdots, m \mid \gamma_i > 0\} \neq \varnothing$, 对任意的 $\varepsilon > 0$, 得

$$x = \sum_{i=1}^{m}\mu_i a_i = \sum_{i=1}^{m}\mu_i a_i - \varepsilon\left(\sum_{i=1}^{m}\gamma_i a_i\right) = \sum_{i=1}^{m}(\mu_i - \varepsilon\gamma_i)a_i.$$

令 $\varepsilon := \min\left\{\dfrac{\mu_i}{\gamma_i}\middle| i \in I\right\} = \dfrac{\mu_{i_0}}{\gamma_{i_0}}$, 其中 $i_0 \in I$, 记 $\beta_i := \mu_i - \varepsilon\gamma_i$, $i = 1, \cdots, m$, 则有关系

$$x = \sum_{i=1}^{m}\beta_i a_i,$$

其中 $\beta_{i_0} = 0$, $\beta_i \geqslant 0$, $i = 1, \cdots, m$.

继续这个过程, 经有限多步后把 x 表示为线性无关元素 $\{a_j | j \in J\}$ 的一个正的线性组合, 其中 $J \subset \{1, \cdots, m\}$. 因此指标集 J 不能包含多于 n 个元素. □

下面著名的结果称为 Carathéodory 定理. 它把集合的凸包与空间的维数联系起来, 参见图 3.2 的说明.

定理 3.6 设 Ω 是 $\mathbb{R}^n$ 的非空子集, 则每个 $x \in \operatorname{co}\Omega$ 可表为集合 Ω 中不超过 $n+1$ 个元素的凸组合.

证明 记 $B := \{1\} \times \Omega \subset \mathbb{R}^{n+1}$, 易见 $\operatorname{co}B = \{1\} \times \operatorname{co}\Omega$, 且 $\operatorname{co}B \subset K_B$. 现任取 $x \in \operatorname{co}\Omega$, 则 $(1, x) \in \operatorname{co}B$ 且根据命题 3.5, 找到 $\lambda_i \geqslant 0$, $(1, a_i) \in B$, $i = 0, \cdots, m$ 且 $m \leqslant n$ 使得

$$(1, x) = \sum_{i=0}^{m}\lambda_i(1, a_i).$$

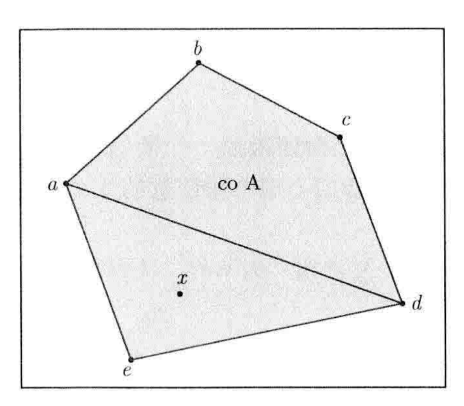


图 3.2 Carathéodory 定理

这蕴涵着 $x = \sum_{i=0}^{m} \lambda_i a_i$, 其中 $\sum_{i=0}^{m} \lambda_i = 1$, $\lambda_i \geqslant 0$, $m \leqslant n$. □

下面有用的断言是定理 3.6 的一个直接结果.

推论 3.7 假设 $\mathbb{R}^n$ 的子集 Ω 是紧的, 则集合 $\mathrm{co}\,\Omega$ 也是紧的.

下面给出两个命题, 它们可使我们得到另外一个著名的结果, 称为 Farkas 引理, 具体内容见定理 3.10.

命题 3.8 设 $a_1, \cdots, a_m$ 是 $\mathbb{R}^n$ 中线性相关的元素, 其中 $a_i \neq 0$, $i = 1, \cdots, m$. 定义集合

$$\Omega := \{a_1, \cdots, a_m\}, \quad \Omega_i := \Omega \setminus \{a_i\}, \quad i = 1, \cdots, m,$$

则由这些集合生成的凸锥可表为

$$K_\Omega = \bigcup_{i=1}^{m} K_{\Omega_i}. \tag{3.3}$$

证明 显然, $\bigcup_{i=1}^{m} K_{\Omega_i} \subset K_\Omega$. 为证相反的包含关系, 取 $x \in K_\Omega$, 则有

$$x = \sum_{i=1}^{m} \mu_i a_i, \quad 其中 \ \ \mu_i \geqslant 0, i = 1, \cdots, m.$$

如果 $x = 0$, 则显然 $x = \bigcup_{i=1}^{m} K_{\Omega_i}$. 否则, a_i 的线性相关性可找到不全为 0 的 $\gamma_i \in \mathbb{R}$ 使得

$$\sum_{i=1}^{m} \gamma_i a_i = 0.$$

与命题 3.5 的证明相同, 把 x 表为

$$x = \sum_{1 \leqslant i \leqslant m,\, i \neq i_0} \lambda_i a_i, \quad 对某 \ \ i_0 \in \{1, \cdots, m\}.$$

这表明 $x \in K_{\Omega_i}$, 因此证明了等式 (3.3) 成立. □

命题 3.9 设 $a_1, \cdots, a_m$ 是 $\mathbb{R}^n$ 的元素, 则由集合 $\Omega := \{a_1, \cdots, a_m\}$ 生成的凸锥 K_Ω 在 $\mathbb{R}^n$ 中是闭的.

证明 先假设元素 $a_1, \cdots, a_m$ 是线性无关的, 并任取收敛于 x 的序列 $\{x_k\} \subset K_\Omega$. 根据 K_Ω 的构造, 找到数 $\alpha_{ki} \geqslant 0, i = 1, \cdots, m, \forall k \in \mathbb{N}$ 使得

$$x_k = \sum_{i=1}^{m} \alpha_{ki} a_i.$$

令 $\alpha_k := (\alpha_{k1}, \cdots, \alpha_{km}) \in \mathbb{R}^m$, 用反证法, 易验证序列 $\{\alpha_k\}$ 是有界的. 不失一般性可假设 $\alpha_k \to (\beta_1, \cdots, \beta_m)(k \to \infty)$ 且 $\alpha_i \geqslant 0, i = 1, \cdots, m$. 因此通过取极限得 $x_k \to x = \sum\limits_{i=1}^{m} \beta_i a_i \in K_\Omega$, 这就证明了 K_Ω 的闭性. □

下面考虑任意元素 $a_1, \cdots, a_m$, 不失一般性, 假设 $a_i \neq 0, i = 1, \cdots, m$. 根据上述证明, 余下需验证当 $a_1, \cdots, a_m$ 是线性相关的情形. 于是根据命题 3.8 有锥表达式 (3.3), 其中每个集合 Ω_i 包含 $m-1$ 个元素. 如果至少一个集合 Ω_i 由线性相关元素构成, 可以用类似的方法通过 $m-2$ 个元素的集合表示相应的锥 K_{Ω_i}. 经有限多步后得

$$K_\Omega = \bigcup_{j=1}^{p} K_{C_j},$$

其中每个集合 $C_j \subset \Omega$ 只包含线性无关的元素, 因此锥 K_{C_j} 是闭的. 于是所考虑的锥 K_Ω 也是闭的. □

现在我们可以得到由 Gyula Farkas 于 1894 年建立的凸分析和最优化的重要结果.

定理 3.10 设 $\Omega := \{a_1, \cdots, a_m\}$, 其中 $a_1, \cdots, a_m \in \mathbb{R}^n$, 设 $b \in \mathbb{R}^n$. 下列断言是等价的:

(i) $b \in K_\Omega$.

(ii) 对任意的 $x \in \mathbb{R}^n$, 有蕴涵关系

$$[\langle a_i, x\rangle \leqslant 0, i = 1, \cdots, m] \Rightarrow [\langle b, x\rangle \leqslant 0].$$

证明 为证 (i)⇒ (ii), 取 $b \in K_\Omega$ 并找到 $\lambda_i \geqslant 0, i = 1, \cdots, m$ 使得

$$b = \sum_{i=1}^{m} \lambda_i a_i,$$

则对给定的 $x \in \mathbb{R}^n$, 不等式 $\langle a_i, x\rangle \leqslant 0, i = 1, \cdots, m$ 蕴涵着

$$\langle b, x\rangle = \sum_{i=1}^{m} \lambda_i \langle a_i, x\rangle \leqslant 0,$$

即 (ii) 成立. 为证方向的蕴涵关系, 假设 $b \notin K_\Omega$, 证明此时 (ii) 不成立. 事实上, 由于集合 K_Ω 是闭凸的, 命题 2.1 的严格分离结果使得存在 $\bar{x} \in \mathbb{R}^n$ 满足

$$\sup\{\langle u, \bar{x}\rangle \mid u \in K_\Omega\} < \langle b, \bar{x}\rangle.$$

由于 K_Ω 是锥, 有 $0 = \langle 0, \bar{x}\rangle < \langle b, \bar{x}\rangle$, $tu \in K_\Omega$, $\forall t > 0$, $u \in K_\Omega$. 因此

$$t\sup\{\langle u, \bar{x}\rangle \mid u \in K_\Omega\} < \langle b, \bar{x}\rangle, \quad \forall t > 0.$$

用 t 去除这个不等式的两边, 并令 $t \to \infty$ 得

$$\sup\{\langle u, \bar{x}\rangle \mid u \in K_\Omega\} \leqslant 0.$$

因此 $\langle a_i, \bar{x}\rangle \leqslant 0$, $i = 1, \cdots, m$, 而 $\langle b, \bar{x}\rangle > 0$. 这是一个矛盾, 从而完成了定理的证明. □

3.3 Radon 定理和 Helly 定理

本节给出由 Edward Helly 于 1913 年发现的凸几何的一个著名定理. 从下面简单的引理开始.

引理 3.11 考虑 $\mathbb{R}^n$ 中任意元素 $w_1, \cdots, w_m$, 其中 $m \geqslant n+2$, 则这些元素是仿射相关的.

证明 在 $\mathbb{R}^n$ 中构造元素 $w_2 - w_1, \cdots, w_m - w_1$, 则这些元素是线性相关的由于 $m - 1 > n$. 因此根据命题 1.63 这些元素 $w_1, \cdots, w_m$ 是仿射相关的. □

下面给出由 Johann Radon 于 1923 年建立的一个定理, 并被他用于证明 Helly 定理.

定理 3.12 考虑 $\mathbb{R}^n$ 中的元素 $w_1, \cdots, w_m$, 其中 $m \geqslant n+2$. 设 $I := \{1, \cdots, m\}$, 则存在 I 的两个非空互不相交的子集 I_1 和 I_2 使得 $I = I_1 \cup I_2$ 且对 $\Omega_1 := \{w_i | i \in I_1\}$, $\Omega_2 := \{w_i | i \in I_2\}$, 有

$$\operatorname{co}\Omega_1 \cap \operatorname{co}\Omega_2 \neq \varnothing.$$

证明 由于 $m \geqslant n+2$, 由引理 3.11 得 $w_1, \cdots, w_n$ 是仿射相关的. 因此存在不全为 0 的实数 $\lambda_1, \cdots, \lambda_m$ 使得

$$\sum_{i=1}^{m} \lambda_i w_i = 0, \quad \sum_{i=1}^{m} \lambda_i = 0.$$

考虑指标集 $I_1 := \{i = 1, \cdots, m \mid \lambda_i \geqslant 0\}$, $I_2 := \{i = 1, \cdots, m \mid \lambda_i < 0\}$. 由于 $\sum_{i=1}^{m} \lambda_i = 0$, 集合 I_1 与 I_2 都是非空的且 $\sum_{i \in I_1} \lambda_i = -\sum_{i \in I_2} \lambda_i$. 记 $\lambda := \sum_{i \in I_1} \lambda_i$, 则有等式

$$\sum_{i \in I_1} \lambda_i w_i = -\sum_{i \in I_2} \lambda_i w_i \quad \text{和} \quad \sum_{i \in I_1} \frac{\lambda_i}{\lambda} w_i = \sum_{i \in I_2} \frac{-\lambda_i}{\lambda} w_i.$$

定义 $\Omega_1 := \{\omega_i | i \in I_1\}, \Omega_2 := \{\omega_i | i \in I_2\}$, 则 $\operatorname{co} \Omega_1 \cap \operatorname{co} \Omega_2 \neq \varnothing$, 这是因为

$$\sum_{i \in I_1} \frac{\lambda_i}{\lambda} w_i = \sum_{i \in I_2} \frac{-\lambda_i}{\lambda} w_i \in \operatorname{co} \Omega_1 \cap \operatorname{co} \Omega_2.$$

这就完成了 Radon 定理的证明. □

现在我们已经准备好阐述和证明重要的 Helly 定理.

定理 3.13 设 $\mathcal{O} := \{\Omega_1, \cdots, \Omega_m\}$ 是 $\mathbb{R}^n$ 中的凸集族, 其中 $m \geqslant n+1$. 假设 $\mathcal{O}$ 中任意 $n+1$ 个集合构成的子族的交是非空的, 那么 $\bigcap_{i=1}^{m} \Omega_i \neq \varnothing$.

证明 相对于 $\mathcal{O}$ 中集合的个数作数学归纳法. 如果 $m = n+1$, 定理的结论是显然的. 假设对 $\mathbb{R}^n$ 中任意 m 个凸集族, 其中 $m \geqslant n+1$, 结论成立. 设 $\{\Omega_1, \cdots, \Omega_m, \Omega_{m+1}\}$ 是 $\mathbb{R}^n$ 中的凸集族使得任意 $n+1$ 个集合构成的子族的交是非空的. 对 $i = 1, \cdots, m+1$, 定义

$$\Theta_i := \bigcap_{1 \leqslant j \leqslant m+1, j \neq i} \Omega_j,$$

并注意到 $\Theta_i \subset \Omega_j, \forall j \neq i, i, j = 1, \cdots, m+1$. 由数学归纳法假设得 $\Theta_i \neq \varnothing$, 因此可取 $w_i \in \Theta_i, i = 1, \cdots, m+1$. 对元素 $w_1, \cdots, w_{m+1}$ 应用上面的 Radon 定理, 可找到两个非空, 互不相交的 $I := \{1, \cdots, m+1\}$ 的子集指标集 I_1 和 I_2 且 $I = I_1 \cup I_2$ 使得对 $W_1 := \{w_i | i \in I_1\}$, $W_2 := \{w_i | i \in I_2\}$ 有 $\operatorname{co} W_1 \cap \operatorname{co} W_2 \neq \varnothing$. 因此取元素 $w \in \operatorname{co} W_1 \cap \operatorname{co} W_2$. 最后证明 $w \in \bigcap_{i=1}^{m+1} \Omega_i$ 从而得结论成立. 由于对 $i \in I_1, j \in I_2, i \neq j$, 因此 $\Theta_i \subset \Omega_j, \forall i \in I_1, j \in I_2$. 任取 $i \in \{1, \cdots, m+1\}$, 考虑 $i \in I_1$ 的情形, 则对任意的 $j \in I_2$, $\omega_j \in \Theta_j \subset \Omega_i$. 因此 $w \in \operatorname{co} W_2 = \operatorname{co}\{w_j | j \in I_2\} \subset \Omega_i$ (由于 Ω_i 的凸性). 这表明对任意的 $i \in I_1$, $w \in \Omega_i$. 同理, 可以证明对任意的 $i \in I_2, w \in \Omega_i$, 因此完成了定理的证明. □

3.4 凸集的切锥

在第 2 章研究了凸集的法锥, 并知道从其定义到重要性质都与凸分离有关. 本节独立地定义凸集的切锥, 然而注意到它也可以通过所考虑集合法锥的对偶来得到.

先考虑一般的锥. 给定非空凸锥 $C \subset \mathbb{R}^n$, C 的极化是

$$C^\circ := \{v \in \mathbb{R}^n | \langle v, c\rangle \leqslant 0, \forall\, c \in C\}.$$

命题 3.14　对任意非空凸锥 $C \subset \mathbb{R}^n$, 有

$$(C^\circ)^\circ = \mathrm{cl}\, C.$$

证明　包含关系 $C \subset (C^\circ)^\circ$ 是定义直接可得, 因此 $\mathrm{cl}\, C \subset (C^\circ)^\circ$, 由于 $(C^\circ)^\circ$ 是闭集. 为证反向包含关系, 任取 $c \subset (C^\circ)^\circ$, 假设 $c \notin \mathrm{cl}\, C$. 由于 C 是锥, 利用命题 2.1 的凸分离, 找到 $v \neq 0$ 使得

$$\langle v, x\rangle \leqslant 0, \quad \forall x \in C \quad 且 \quad \langle v, c\rangle > 0,$$

因此 $v \in C^\circ$, 从而与 $c \in (C^\circ)^\circ$ 矛盾. □

定义 3.15　设 Ω 是非空凸集, Ω 在 $\bar{x} \in \Omega$ 处的切锥定义为

$$T(\bar{x}; \Omega) := \mathrm{cl}\, K_{\{\Omega - \bar{x}\}} = \mathrm{cl}\, \mathbb{R}^+(\Omega - \bar{x}).$$

现在我们导出凸集的法锥和切锥之间的对偶关系; 见图 3.3.

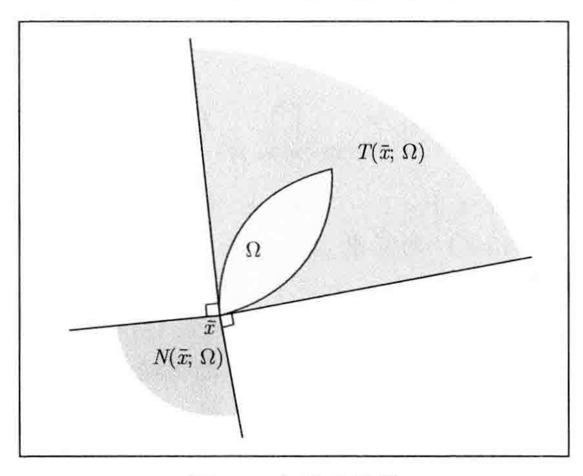


图 3.3　切锥和法锥

定理 3.16　设 Ω 是凸集, $\bar{x} \in \Omega$, 则有等式

$$N(\bar{x}; \Omega) = [T(\bar{x}; \Omega)]^\circ, \quad T(\bar{x}; \Omega) = [N(\bar{x}; \Omega)]^\circ. \tag{3.4}$$

证明　为证第一个包含关系 $N(\bar{x}; \Omega) \subset [T(\bar{x}; \Omega)]^\circ$, 任取 $v \in N(\bar{x}; \Omega)$, 由定义得 $\langle v, x - \bar{x}\rangle \leqslant 0, \forall\, x \in \Omega$. 这蕴涵着

$$\langle v, t(x - \bar{x})\rangle \leqslant 0, \quad \forall t \geqslant 0, \quad x \in \Omega.$$

取定 $w \in T(\bar{x};\Omega)$, 可找到 $t_k \geqslant 0, x_k \in \Omega$ 且 $t_k(x_k - \bar{x}) \to w(k \to \infty)$, 则

$$\langle v, w\rangle = \lim_{k\to\infty} \langle v, t_k(x_k - \bar{x})\rangle \leqslant 0,$$

因此 $v \in [T(\bar{x};\Omega)]^0$.

为证式 (3.4) 中第一个关系的反向包含关系成立, 任取 $v \in [T(\bar{x};\Omega)]^\circ$, 由于当 $w \in T(\bar{x};\Omega)$ 时 $\langle v, w\rangle$ 和当 $x \in \Omega$ 时 $x - \bar{x} \in T(\bar{x};\Omega)$, 因为 $x \in \Omega$, 推得

$$\langle v, x - \bar{x}\rangle \leqslant 0, \quad \forall x \in \Omega.$$

因此 $w \in N(\bar{x};\Omega)$, 这就完成了式 (3.4) 中第一个等式的证明. 式 (3.4) 中第二个等式由第一个等式可得, 这是因为由命题 3.14 知

$$[N(\bar{x};\Omega)]^\circ = [T(\bar{x};\Omega)]^{\circ\circ} = \operatorname{cl} T(\bar{x};\Omega),$$

而根据定义 3.15, $T(\bar{x};\Omega)$ 是闭的. □

接下来的结果部分基于 Farkas 引理 (定理 3.10), 给出线性不等式组的切锥和法锥的有效表示.

定理 3.17 设 $a_i \in \mathbb{R}^n$, 设 $b_i \in \mathbb{R}, i = 1, \cdots, m$, 并设

$$\Omega := \{x \in \mathbb{R}^n | \langle a_i, x\rangle \leqslant b_i\}.$$

假设 $\bar{x} \in \Omega$, 并定义活跃指标集 $I(\bar{x}) := \{i = 1, \cdots, m | \langle a_i, \bar{x}\rangle = b_i\}$, 则有关系式

$$T(\bar{x};\Omega) = \{v \in \mathbb{R}^n | \langle a_i, v\rangle \leqslant 0, \forall i \in I(\bar{x})\}, \tag{3.5}$$

$$N(\bar{x};\Omega) = \operatorname{cone}\{a_i | i \in I(\bar{x})\}. \tag{3.6}$$

其中按惯例置 $\operatorname{cone}\phi = \{0\}$.

证明 为证式 (3.5). 考虑凸锥 $K := \{v \in \mathbb{R}^n | \langle a_i, v\rangle \leqslant 0, \forall_i \in I(\bar{x})\}$.

易见对任意的 $x \in \Omega$, $i \in I(\bar{x})$, $t \geqslant 0$, 有

$$\langle a_i, t(x - \bar{x})\rangle = t(\langle a_i, x\rangle - \langle a_i, \bar{x}\rangle) \leqslant t(b_i - b_i) = 0,$$

因此 $\mathbb{R}^+(\Omega - \bar{x}) \subset K$. 于是由 K 的闭性得

$$T(\bar{x};\Omega) = \text{cl}1, \mathbb{R}^+(\Omega - \bar{x}) \subset K.$$

为证式 (3.5) 中反向的包含关系成立, 注意到 $\langle a_i, \bar{x}\rangle < b_i$, $i \notin I(\bar{x})$. 利用这个结果并任取 $v \in K$, 有

$$\langle a_i, \bar{x}+tv\rangle \leqslant b_i \text{ 对 } \forall i \notin I(\bar{x}) \text{ 和任意小的 } t>0 \text{ 成立.}$$

由于 $\langle a_i, v\rangle \leqslant 0, \forall i \in I(\bar{x})$, 因此 $\langle a_i, \bar{x}+tv\rangle \leqslant b_i, i=1,\cdots,m$, 从而有

$$\bar{x}+tv \in \Omega, \quad \text{或} \quad v \in \frac{1}{t}(\Omega-\bar{x}) \subset T(\bar{x};\Omega).$$

这就证明了切锥表达式 (3.5).

下证法锥表达式 (3.6). 式 (3.4) 中的第一个等式保证

$$w \in N(\bar{x};\Omega) \quad \text{当且仅当} \quad \langle w, v\rangle \leqslant 0, \quad \forall v \in T(\bar{x};\Omega),$$

则由式 (3.5) 得 $w \in N(\bar{x};\Omega)$ 当且仅当对任意的 $v \in \mathbb{R}^n$, 有

$$[\langle a_i, v\rangle \leqslant 0, \forall i \in I(\bar{x})] \Rightarrow [\langle w, v\rangle \leqslant 0].$$

由定理 3.10 中的 Farkas 引理知 $w \in N(\bar{x};\Omega)$ 当且仅当 $w \in \operatorname{cone}\{a_i | i \in I(\bar{x})\}$, 因此式 (3.6) 成立. □

3.5 中值定理

关于可微函数的中值定理是经典实分析的重要结果之一. 现在对一般凸函数导出其次微分版本.

定理 3.18 设 $f:\mathbb{R}\to\overline{\mathbb{R}}$ 是凸的, $a<b$, 并设区间 $[a,b]$ 包含于 f 的定义域, 其中 f 的定义域是 $\mathbb{R}$ 中的开区间, 则存在 $c\in(a,b)$ 使得

$$\frac{f(b)-f(a)}{b-a} \in \partial f(c). \tag{3.7}$$

证明 定义实值函数 $g:[a,b]\to\mathbb{R}$ 为

$$g(x) := f(x) - \left[\frac{f(b)-f(a)}{b-a}(x-a)+f(a)\right],$$

则 $g(a)=g(b)$. 根据 g 在它的开定义域的凸性, 得有其连续性, 由此 g 在某点 $c\in(a,b)$ 处有局部极小值. 函数

$$h(x) := -\left[\frac{f(b)-f(a)}{b-a}(x-a)+f(a)\right]$$

显然在 c 点是可微的, 因此

$$\partial h(c) = \{h'(c)\} = \left\{-\frac{f(b)-f(a)}{b-a}\right\}.$$

由第 2 章的次微分 Fermat 法则和次微分的和法则得

$$0 \in \partial g(c) = \partial f(c) - \left\{ \frac{f(b) - f(a)}{b - a} \right\},$$

这蕴涵着式 (3.7) 成立, 从而定理的证明. □

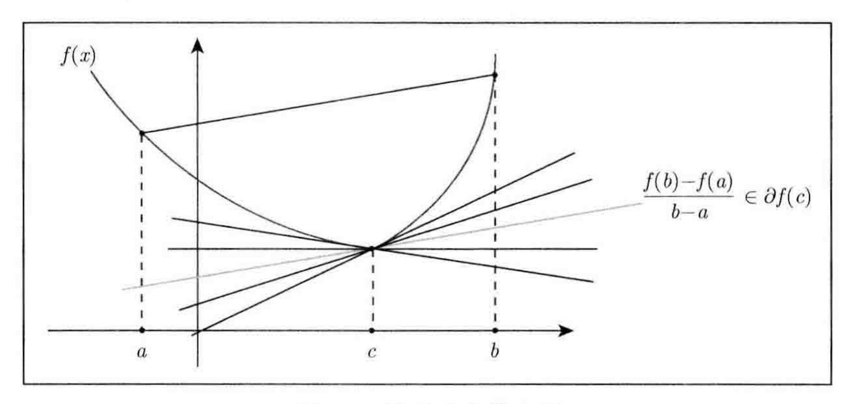


图 3.4 次微分中值定理

为进一步得到 $\mathbb{R}^n$ 上凸函数相应的中值定理 (定理 3.20), 首先给出下面的引理.

引理 3.19 设 $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸函数, D 是 f 的定义域, 它是 $\mathbb{R}^n$ 的开子集. 给定 $a, b \in D$, 且 $a \neq b$, 定义函数 $\varphi: \mathbb{R} \to \overline{\mathbb{R}}$ 为

$$\varphi(t) := f(tb + (1-t)a), \quad t \in \mathbb{R}, \tag{3.8}$$

则对任意的数 $t_0 \in (0, 1)$, 有

$$\partial\varphi(t_0) = \{\langle v, b - a\rangle | v \in \partial f(c_0)\}, \quad 其中 \ \ c_0 := t_0 a + (1 - t_0)b.$$

证明 分别定义向量值函数 $A: \mathbb{R} \to \mathbb{R}^n$ 和 $B: \mathbb{R} \to \mathbb{R}^n$ 为

$$A(t) := (b - a)t, \quad B(t) := tb + (1-t)a = t(b-a) + a = A(t) + a, \quad t \in \mathbb{R}.$$

显然 $\varphi(t) = (f \circ B)(t)$, A 的伴随映射是 $A^*v = \langle v, b - a\rangle, v \in \mathbb{R}^n$. 根据定理 2.51 中的次微分链式法则, 有

$$\partial\varphi(t_0) = A^*\partial f(c_0) = \{\langle v, b - a\rangle | v \in \partial f(c_0)\},$$

这就证明了引理中所断言的结果. □

定理 3.20 设 $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ 是引理 3.19 中的凸函数, 则存在元素 $c \in (a, b)$ 使得

$$f(b) - f(a) \in \langle \partial f(c), b - a\rangle := \{\langle v, b - a\rangle | v \in \partial f(c)\}.$$

证明　考虑式 (3.8) 中的函数 $\varphi : \mathbb{R} \to \overline{\mathbb{R}}$, 则有 $\varphi(0) = f(a), \varphi(1) = f(b)$. 对 φ 首先应用定理 3.18, 然后应用引理 3.19, 可找到 $t_0 \in (0,1)$ 使得

$$f(b) - f(a) = \varphi(1) - \varphi(0) \in \partial\varphi(t_0) = \{\langle v, b-a\rangle | v \in \partial f(c)\},$$

其中 $c = t_0 a + (1-t_0)b$. 这就证明了我们的结论. □

3.6　地　平　锥

本节研究无界凸集在 ∞ 处的性质.

定义 3.21　给定 $\mathbb{R}^n$ 的非空闭凸子集 F, 给定点 $x \in F$, F 在点 x 处的地平锥如下

$$F_\infty(x) := \{d \in \mathbb{R}^n | x + td \in F, \forall t > 0\}.$$

注意到地平锥在文献中也称为 F 在所论点处的渐近锥. $F_\infty(x)$ 的另一等价定义由

$$F_\infty(x) = \bigcap_{t>0} \frac{F-x}{t}$$

给出, 这显然蕴涵着 $F_\infty(x)$ 是 $\mathbb{R}^n$ 中的闭凸锥. 接下来的命题表明对任意的 $x \in F, F_\infty(x)$ 是相同的, 因此可简记为 F_∞.

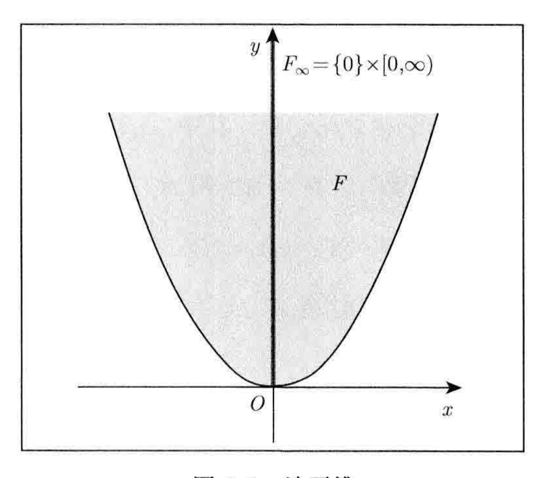


图 3.5　地平锥

命题 3.22　设 F 是 $\mathbb{R}^n$ 的非空闭凸子集, 则对任意的 $x_1, x_2 \in F$, 有等式 $F_\infty(x_1) = F_\infty(x_2)$.

证明 只需证明对任意的 $x_1, x_2 \in F$, $F_\infty(x_1) \subset F_\infty(x_2)$. 任取方向 $d \in F_\infty(x_1)$, 任意数 $t > 0$, 证明 $x_2 + td \in F$. 考虑序列

$$x_k := \frac{1}{k}(x_1 + ktd) + \left(1 - \frac{1}{k}\right) x_2, \quad k \in \mathbb{N},$$

则对任意的 k, 因为 $d \in F_\infty(x_1)$ 且 F 是凸的, 所以 $x_k \in F$. 我们还有 $x_k \to x_2 + td$, 因此 $x_2 + td \in F$(由于 F 是闭的). 故 $d \in F_\infty(x_2)$. □

推论 3.23 设 F 是 $\mathbb{R}^n$ 包含原点的闭凸子集, 则

$$F_\infty = \bigcap_{t>0} tF.$$

证明 由 F_∞ 的构造和命题 3.22 得

$$F_\infty = F_\infty(0) = \bigcup_{t>0} \frac{F - 0}{t} = \bigcap_{t>0} \frac{1}{t} F = \bigcap_{t>0} tF,$$

这就推论成立. □

命题 3.24 设 F 是 $\mathbb{R}^n$ 的非空闭凸子集, 对给定的元素 $d \in \mathbb{R}^n$, 下面的断言是等价的:

(i) $d \in F_\infty$.

(ii) 存在序列 $\{t_k\} \subset [0, \infty), \{f_k\} \subset F$ 使得 $t_k \to 0$, $t_k f_k \to d$.

证明 为证 (i)⇒(ii), 取 $d \in F_\infty$, 取定 $\bar{x} \in F$. 由定义得

$$\bar{x} + kd \in F, \quad \forall k \in \mathbb{N}.$$

对每个 $k \in \mathbb{N}$ 可找到 $f_k \in F$ 使得

$$\bar{x} + kd = f_k, \quad \text{或等价地}, \quad \frac{1}{k}\bar{x} + d = \frac{1}{k} f_k.$$

令 $t_k := \dfrac{1}{k}$, 得 $t_k f_k \to d(k \to \infty)$.

为证 (ii)⇒(i), 假设存在序列 $\{t_k\} \subset [0, \infty)$, $\{f_k\} \subset F$ 使得 $t_k \to 0, t_k f_k \to d$. 取定 $x \in F$, 下面验证

$$x + td \in F, \quad \forall t > 0$$

从而有 $d \in F_\infty$. 事实上, 对任意固定的 $t > 0$, 当 k 充分大时, 有 $0 \leqslant t \cdot t_k < 1$. 因此

$$(1 - t \cdot t_k)x + t \cdot t_k f_k \to x + td(k \to \infty).$$

由 F 的凸性得 $(1 - t \cdot t_k)x + t \cdot t_k f_k \in F$. 由闭性知 $x + td \in F$, 从而得 $d \in F_\infty$. □

在本节最后, 利用地平锥给出集合有界性所期待的刻画.

定理 3.25 设 F 是 $\mathbb{R}^n$ 的非空闭凸子集, 则集合 F 是有界的当且仅当它的地平锥是平凡的, 即 $F_\infty = \{0\}$.

证明 假设 F 是有界的, 任取 $d \in F_\infty$. 根据命题 3.24, 存在序列 $\{t_k\} \subset [0, \infty)$ 且 $t_k \to 0, \{f_k\} \subset F$ 使得 $t_k f_k \to d(k \to \infty)$. 由 F 的有界性得 $t_k f_k \to 0$, 这表明 $d = 0$.

为证相反的蕴涵关系成立. 假设 F 是无界的, 而 $F_\infty = \{0\}$. 则存在序列 $\{x_k\} \subset F$ 且 $\|x_k\| \to \infty$. 于是可构造 (单位) 方向序列

$$d_k := \frac{x_k}{\|x_k\|}, \quad k \in \mathbb{N}.$$

不失一般性, 设 $d_k \to d(k \to \infty)$ 且 $\|d\| = 1$. 任取 $x \in F$, 并注意到对任意的 $t > 0$ 和充分大的 $k \in \mathbb{N}$, 根据 F 的凸性得

$$u_k := \left(1 - \frac{t}{\|x_k\|}\right) x + \frac{t}{\|x_k\|} x_k \in F.$$

由于 $u_k \to x + td$ 以及 F 是闭的, 所以 $x + td \in F$, 所以 $d \in F_\infty$, 矛盾, 这就完成了定理的证明. □

3.7 极小时间函数和 Minkowski 度规

本节及 3.8 节的主要研究对象是一类不太普通的凸函数, 称为极小时间函数. 在凸分析中它们是相对新的, 但在应用中是相当重要的. 它们在选址问题中的一些近期应用将在第 4 章中给出. 本节还讨论凸分析和最优化中公认的极小时间函数与 Minkowski 度规函数之间的关系.

定义 3.26 设 F 是 $\mathbb{R}^n$ 的非空闭凸子集, 设 $\Omega \subset \mathbb{R}^n$ 是非空集合, 不一定是凸的, 则关联于集合 F 和 Ω 的极小时间函数定义为

$$\mathcal{T}_\Omega^F(x) := \inf\{t \geqslant 0 | (x + tF) \cap \Omega \neq \varnothing\}. \tag{3.9}$$

这个名称来自下面的解释: 函数 (3.9) 表明在恒定的动态 F 作用下点 x 到达目标集合 Ω 所需的最小时间; 参见图 3.6. 注意到当 $F = \mathbb{B}$, $\mathbb{R}^n$ 中的闭单位球时, 极小时间函数 (3.9) 即为 2.5 节考虑的距离函数. 当 $F = \{0\}$ 时, 极小时间函数 (3.9) 是 Ω 的指示函数. 如果 $\Omega = \{0\}$, 我们得 $\mathcal{T}_\Omega^F(x) = \rho_F(-x)$, 其中 ρ_F 是本节后面定义的 Minkowski 度规. 极小时间函数 (3.9) 还包括另一类有趣的函数称为方向极小时间函数, 对应于当 F 只含一个非零向量的情形.

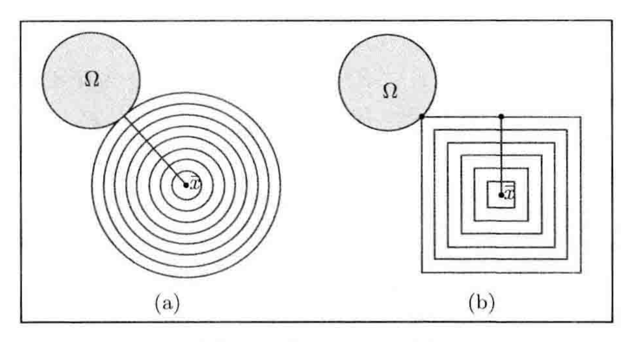


图 3.6 极小时间函数

本节我们将研究用于下面的极小时间函数 (3.9) 的一般性质. 当然这些性质本身也具有独立的意义.

注意到, 如果 $x \in \Omega$, 则$\mathcal{T}_{\Omega}^{F}(x) = 0$. 如果存在 $t \geqslant 0$ 且 $(x + tF) \cap \Omega \neq \varnothing$, 则 $\mathcal{T}_{\Omega}^{F}(x)$ 是有限的. 特别地, 当 $0 \in \operatorname{int} F$ 时, $\mathcal{T}_{\Omega}^{F}(x)$ 是有限的. 如果没有这样的 t 存在, 那么 $\mathcal{T}_{\Omega}^{F}(x) = \infty$. 因此极小时间函数 (3.9) 一般来说是增广实值函数.

式 (3.9) 中的集合 F 和 Ω 不一定是有界的. 下面的定理揭示了当 Ω 是有界的而 F 不一定是有界的情形下 $\mathcal{T}_{\Omega}^{F}(x)$ 的性质. 对 F 是有界的, 而 Ω 不一定是有界的情形下相关结果的表达和证明留作练习.

定理 3.27 设 F 是 $\mathbb{R}^n$ 的非空闭凸子集, 设 $\Omega \subset \mathbb{R}^n$ 是有界非空闭集, 则

$$\mathcal{T}_{\Omega}^{F}(x) = 0 \text{ 蕴涵着 } x \in \Omega - F_{\infty}.$$

如果再加上假设 $0 \in F$, 则反向的结论成立.

证明 假设 $\mathcal{T}_{\Omega}^{F}(x) = 0$, 找到 $t_k \to 0, t_k \geqslant 0$, 使得

$$(x + t_k F) \cap \Omega \neq \varnothing, \quad k \in \mathbb{N}.$$

因此存在 $f_k \in F$ 使得 $x + t_k f_k = q_k \in \Omega, \forall k$. 根据 Ω 的有界性和闭性, 可选择收敛于某 $q \in \Omega$ 的 $\{q_k\}$ 的子序列, 则沿此序列有 $t_k f_k \to q - x$. 由命题 3.24 得 $q - x \in F_{\infty}$, 这就证明了 $x \in \Omega - F_{\infty}$.

为证反向的蕴涵关系成立, 任取 $x \in \Omega - F_{\infty}$, 则可找到 $q \in \Omega, d \in F_{\infty}$ 使得 $x = q - d$. 由于 $0 \in F, d \in F_{\infty}$, 由推论 3.23 得 $k(q - x) = kd \in F, \ \forall k \in \mathbb{N}$. 这表明

$$q - x \in \frac{1}{k}F,$$

因此 $\left(x + \dfrac{1}{k}F\right) \cap \Omega \neq \varnothing, k \in \mathbb{N}$.

上面第二个式子保证 $0 \leqslant \mathcal{T}_{\Omega}^{F}(x) \leqslant \dfrac{1}{k}, \forall k \in \mathbb{N}$. 它成立仅当 $\mathcal{T}_{\Omega}^{F}(x) = 0$, 这就完成了定理的证明. □

我们用下面的例子来解释说明定理 3.27 的结果.

例 3.28 设 $F = \mathbb{R} \times [-1,1] \subset \mathbb{R}^2$, 设 Ω 是以 $(1,0)$ 为中心, 半径为 $r=1$ 的圆盘. 利用推论 3.23 中 F_∞ 的表示, 有 $F_\infty = \mathbb{R} \times \{0\}, \Omega - F_\infty = \mathbb{R} \times [-1,1]$. 因此 $\mathcal{T}_\Omega^F(x) = 0$ 当且仅当 $x \in \mathbb{R} \times [-1,1]$.

接下来的结果刻画了极小时间函数 (3.9) 的凸性.

命题 3.29 设 F 是 $\mathbb{R}^n$ 的非空闭凸子集. 如果 $\Omega \subset \mathbb{R}^n$ 是非空凸的, 则极小时间函数 (3.9) 是凸的. 进一步, 如果 F 是有界的, Ω 是闭的, 则反向的结论也成立.

证明 任取 $x_1, x_2 \in \operatorname{dom} \mathcal{T}_\Omega^F$, 设 $\lambda \in (0,1)$. 首先在集合 F 和 Ω 都是凸的条件下证明

$$\mathcal{T}_\Omega^F(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda \mathcal{T}_\Omega^F(x_1) + (1-\lambda)\mathcal{T}_\Omega^F(x_2). \tag{3.10}$$

记 $\gamma_i := \mathcal{T}_\Omega^F(x_i), i = 1,2$, 任给 $\varepsilon > 0$, 选取数 t_i 使得

$$\gamma_i \leqslant t_i < \gamma_i + \varepsilon, \quad (x_i + t_i F) \cap \Omega \neq \varnothing, \quad i = 1,2.$$

由 F 和 Ω 的凸性得

$$[\lambda x_1 + (1-\lambda)x_2 + (\lambda t_1 + (1-\lambda)t_2)F] \cap \Omega \neq \varnothing,$$

这进一步蕴涵着估计

$$\mathcal{T}_\Omega^F(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda t_1 + (1-\lambda)t_2 \leqslant \lambda \mathcal{T}_\Omega^F(x_1) + (1-\lambda)\mathcal{T}_\Omega^F(x_2) + \varepsilon.$$

由于 ε 是任意的, 则式 (3.10) 成立, 从而证明了 $\mathcal{T}_\Omega^F$ 的凸性.

反之, 假设在 F 是有界的, Ω 是闭的条件下 $\mathcal{T}_\Omega^F$ 是凸的, 则水平集

$$\{x \in \mathbb{R}^n | \mathcal{T}_\Omega^F(x) \leqslant 0\}$$

一定是凸的. 易证该集合即为目标集 Ω, 这就证明了 Ω 的凸性. □

定义 3.30 设 F 是 $\mathbb{R}^n$ 的非空闭凸子集. 关联于集合 F 的 Minkowski 度规定义为

$$\rho_F(x) := \inf\{t \geqslant 0 \mid x \in tF\}, \quad x \in \mathbb{R}^n. \tag{3.11}$$

下面的定理概括了 Minkowski 度规函数的主要性质.

定理 3.31 设 F 是包含原点的闭凸集合, 则 Minkowski 度规 ρ_F 是增广实值函数, 而且是正齐次和次可加的. 进一步, $\rho_F(x) = 0$ 当且仅当 $x \in F_\infty$.

证明 先证 ρ_F 是次可加的, 即

$$\rho_F(x_1 + x_2) \leqslant \rho_F(x_1) + \rho_F(x_2), \quad \forall x_1, x_2 \in \mathbb{R}^n. \tag{3.12}$$

如果式 (3.12) 的右边等于 ∞, 即有 $x_1 \notin \operatorname{dom}\rho_F$ 或 $x_2 \notin \operatorname{dom}\rho_F$, 式 (3.12) 显然成立, 现假设 $x_1, x_2 \in \operatorname{dom}\rho_F$, 取定 $\varepsilon > 0$, 则存在数 $t_1, t_2 \geqslant 0$ 使得

$$\rho_F(x_1) \leqslant t_1 < \rho_F(x_1) + \varepsilon, \quad \rho_F(x_2) \leqslant t_2 < \rho_F(x_2) + \varepsilon \text{ 且 } x_1 \in t_1 F, x_2 \in t_2 F.$$

由 F 的凸性得 $x_1 + x_2 \in t_1F + t_2F \subset (t_1 + t_2)F$, 因此

$$\rho_F(x_1 + x_2) \leqslant t_1 + t_2 < \rho_F(x_1) + \rho_F(x_2) + 2\varepsilon,$$

由 $\varepsilon > 0$ 的任意性得式 (3.12) 成立. 为验证 ρ_F 的正齐次性质, 任取 $\alpha > 0$, 则

$$\begin{aligned}\rho_F(\alpha x) &= \inf\{t \geqslant 0 | \alpha x \in tF\} = \inf\left\{t \geqslant 0 \middle| x \in \frac{t}{\alpha}F\right\} \\ &= \alpha \inf\left\{\frac{t}{\alpha} \geqslant 0 \middle| x \in \frac{t}{\alpha}F\right\} = \alpha\rho_F(x).\end{aligned}$$

余下证明定理的最后一个论断成立. 注意到 $\rho_F(x) = \mathcal{T}_\Omega^{-F}(x)$ 是当 $\Omega = \{0\}$ 时极小时间函数的特殊情形. 由于 $0 \in F$, 由定理 3.27 知 $\rho_F(x) = 0$ 当且仅当 $x \in \Omega - (-F)_\infty$, 这就完成了证明. □

为简单起见, 称 $\varphi : \mathbb{R}^n \to \mathbb{R}$ 在 $\mathbb{R}^n$ 上是 ℓ-Lipschitz 连续的, 如果它在 $\mathbb{R}^n$ 上具有 Lipschitz 常数 $\ell \geqslant 0$ 是 Lipschitz 连续的. 下面的结果计算了 $\mathbb{R}^n$ 上 Minkowski 度规的 Lipschitz 常数 ℓ.

命题 3.32 设 F 是包含原点作为内点的闭凸集, 则 Minkowski 度规 ρ_F 在 $\mathbb{R}^n$ 上是 ℓ-Lipschitz 连续的, 且

$$\ell := \inf\left\{\frac{1}{r}\middle| \mathbb{B}(0;r) \subset F,\ r > 0\right\}. \tag{3.13}$$

特别地, 有 $\rho_F(x) \leqslant \ell\|x\|, \forall x \in \mathbb{R}^n$.

证明 条件 $0 \in \operatorname{int} F$ 保证式 (3.13) 中的常数 $\ell < \infty$. 任取 $r > 0$ 满足 $\mathbb{B}(0;r) \subset F$, 由定义得

$$\rho_F(x) = \inf\{t \geqslant 0 \mid x \in tF\} \leqslant \inf\{t \geqslant 0 | x \in t\mathbb{B}(0;r)\} = \inf\{t \geqslant 0 | \|x\| \leqslant rt\} = \frac{\|x\|}{r},$$

这蕴涵着 $\rho_F(x) \leqslant \ell\|x\|$. 于是由 ρ_F 的次可加性得

$$\rho_F(x) - \rho_F(y) \leqslant \rho_F(x - y) \leqslant \ell\|x - y\|, \quad \forall x, y \in \mathbb{R}^n,$$

因此 ρ_F 在 $\mathbb{R}^n$ 上是 ℓ-Lipschitz 连续的, 而 ℓ 由式 (3.13) 给出. □

下面的结果将极小时间函数与 Minkowski 度规联系起来.

定理 3.33 设 F 是非空闭凸集合, 则对任意的 $\Omega \neq \varnothing$, 有

$$\mathcal{T}_\Omega^F(x) = \inf\{\rho_F(\omega - x) | \omega \in \Omega\}, \quad x \in \mathbb{R}^n. \tag{3.14}$$

证明 首先考虑当 $\mathcal{T}_{\Omega}^{F}(x)=\infty$ 时的情形. 此时有

$$\{t\geqslant 0 \mid (x+tF)\cap\Omega\neq\varnothing\}=\varnothing,$$

这蕴涵着对任意的 $\omega\in\Omega$, 有 $\{t\geqslant 0|\omega-x\in tF\}=\varnothing$. 因此 $\rho_F(\omega-x)=\infty, \forall\omega\in\Omega$, 故式 (3.14) 的右边也是 ∞.

现考虑当 $T_{\Omega}^{F}(x)<\infty$ 时的情形. 任取 $t\geqslant 0$ 使得 $(x+tF)\cap\Omega\neq\varnothing$. 即有 $f\in F, \omega\in\Omega$ 且 $x+tf=\omega$. 于是 $\omega-x\in tF$, 因此 $\rho_F(\omega-x)\leqslant t$, 从而 $\inf_{D\in\Omega}\rho_F(\omega-x)\leqslant t$. 由定义 (3.9) 得

$$\inf_{\omega\in\Omega}\rho_F(\omega-x)\leqslant \mathcal{T}_{\Omega}^{F}(x)<\infty.$$

为证式 (3.14) 中反向不等式成立, 记 $\gamma:=\inf\limits_{\omega\in\Omega}\inf\limits_{\omega\in\Omega}\rho_F(\omega-x)$, 对任意的 $\varepsilon>0$, 找到 $\omega\in\Omega$ 满足 $\rho_F(\omega-x)<\gamma+\varepsilon$. 根据 Minkowski 度规的定义 (3.11), 有 $t\geqslant 0$ 使得 $t<\gamma+\varepsilon$ 且 $\omega-x\in tF$, 即有 $\omega\in x+tF$, 因此

$$\mathcal{T}_{\Omega}^{F}(x)\leqslant t<\gamma+\varepsilon.$$

由 ε 的任意性, 得

$$\mathcal{T}_{\Omega}^{F}(x)\leqslant\gamma=\inf_{\omega\in\Omega}\rho_F(\omega-x),$$

这蕴涵着式 (3.14) 中余下的不等式并完成了证明. □

作为所得结果和命题 3.32 的一个推论, 现在建立极小时间函数的 ℓ-Lipschitz 连续性, 其中 ℓ 由式 (3.13) 给出.

推论 3.34 设 F 是包含原点作为内点的闭凸集合, 则对任意非空集合 Ω, 函数 $\mathcal{T}_{\Omega}^{F}$ 是 ℓ-Lipschitz 的, 具常数式 (3.13).

证明 由 ρ_F 的次可加性和命题 3.32 的结果可得

$$\rho_F(\omega-x)\leqslant\rho_F(\omega-y)+\rho_F(y-x)\leqslant\rho_F(\omega-y)+\ell\|y-x\|,$$

对任意的 $\omega\in\Omega, \forall x,y\in\mathbb{R}^n$ 成立, 则由定理 3.33 知

$$\mathcal{T}_{\Omega}^{F}(x)\leqslant\mathcal{T}_{\Omega}^{F}(y)+\ell\|x-y\|.$$

这蕴涵着 $\mathcal{T}_{\Omega}^{F}$ 的 ℓ-Lipschitz 性质, 因为 x,y 是任意选取的. □

下面通过极小时间函数定义目标集合 Ω 的扩张为

$$\Omega_r:=\{x\in\mathbb{R}^n|\mathcal{T}_{\Omega}^{F}(x)\leqslant r\},\quad r>0. \tag{3.15}$$

命题 3.35 设 F 是 $\mathbb{R}^n$ 的非空闭凸子集, $\Omega \subset \mathbb{R}^n$ 是非空集合, 则对任意非空集合 Ω 和任意的 $x \notin \Omega_r$ 且 $\mathcal{T}_\Omega^F(x) < \infty$, 有

$$\mathcal{T}_\Omega^F(x) = \mathcal{T}_{\Omega_r}^F + r.$$

证明 由 $\Omega \subset \Omega_r$ 得 $\mathcal{T}_{\Omega_r}^F(x) \leqslant \mathcal{T}_\Omega^F(x) < \infty$. 任取 $t \geqslant 0$ 满足

$$(x + tF) \cap \Omega_r \neq \varnothing,$$

找到 $f_1 \in F, u \in \Omega_r$ 使得 $x + tf_1 = u$. 由于 $u \in \Omega_r$, 有 $\mathcal{T}_\Omega^F(u) \leqslant r$. 对任意的 $\varepsilon > 0$, 存在 $s \geqslant 0$ 使得 $s < r + \varepsilon$, $(u + sF) \cap \Omega \neq \varnothing$. 这蕴涵着存在 $\omega \in \Omega$, $f_2 \in F$ 使得 $u + sf_2 = \omega$. 由于 F 是凸的, 有

$$\omega = u + sf_2 = (x + tf_1) + sf_2 \in x + (t + s)F.$$

由于 F 是凸的, 由 $\varepsilon > 0$ 的任意性, 得

$$\mathcal{T}_\Omega^F(x) \leqslant t + s \leqslant t + r + \varepsilon,$$

因此 $\mathcal{T}_\Omega^F(x) \leqslant t + r$. 所以有

$$\mathcal{T}_\Omega^F(x) \leqslant \mathcal{T}_{\Omega_r}^F + r.$$

为证反向的不等式成立, 记 $\gamma := \mathcal{T}_\Omega^F(x)$, 注意到由于 $x \notin \Omega_r$, 所以 $r < \gamma$. 对任意的 $\varepsilon > 0$, 找到 $t \in [0, \infty)$, $f \in F$ 和 $\omega \in \Omega$ 使得 $\gamma \leqslant t < \gamma + \varepsilon$, $x + tf = \omega$. 因此

$$\omega = x + tf = x + (t - r)f + rf \in x + (t - r)f + rF,$$

因此 $\mathcal{T}_\Omega^F(x + (t - r)f) \leqslant r$, 这就证明了 $x + (t - r)f \in \Omega_r$, 且还看到 $x + (t - r)f \in x + (t - r)F$. 于是 $\mathcal{T}_{\Omega_r}^F(x) \leqslant t - r \leqslant \gamma - r + \varepsilon$, 由 $\varepsilon > 0$ 的任意性这就证明了

$$r + \mathcal{T}_{\Omega_r}^F(x) \leqslant \gamma = \mathcal{T}_\Omega^F(x)$$

成立, 从而完成了证明. □

命题 3.36 设 F 是 $\mathbb{R}^n$ 的非空闭凸子集, 则对任意非空集合 Ω, 任意的 $x \in \operatorname{dom} \mathcal{T}_\Omega^F$, 任意的 $t \geqslant 0$, 任意的 $f \in F$, 有

$$\mathcal{T}_\Omega^F(x - tf) \leqslant \mathcal{T}_\Omega^F(x) + t.$$

证明 取定 $\varepsilon > 0$, 选取 $s \geqslant 0$ 使得

$$\mathcal{T}_\Omega^F(x) \leqslant s < \mathcal{T}_\Omega^F(x) + \varepsilon, \quad \text{且}(x + sF) \cap \Omega \neq \varnothing.$$

于是 $(x - tf + tF + sF) \cap \Omega \neq \varnothing$, 因此 $(x - tf + (t + s)F) \cap \Omega \neq \varnothing$. 这表明

$$\mathcal{T}_\Omega^F(x - tf) \leqslant t + s \leqslant t + \mathcal{T}_\Omega^F(x) + \varepsilon,$$

令 $\varepsilon \to 0^+$, 这就蕴涵着命题结论成立. □

3.8 极小时间函数的次梯度

注意到极小时间函数 (3.9) 本质上是非光滑的, 这里我们研究当目标集合 Ω 是凸的情形下它的广义微分性质. 此时 (由于恒定动态 F 总可假设为凸的) 根据命题 3.29 函数 (3.9) 也是凸的. 下面的结果根据 $\bar{x}$ 在集合 $\Omega - F_\infty$ 中的位置, 给出了计算次微分 $\partial\mathcal{T}_\Omega^F(\bar{x})$ 的精确表达式.

定理 3.37　设两个集合 $0 \in F \subset \mathbb{R}^n$ 与 $\Omega \subset \mathbb{R}^n$ 都是非空闭凸的. 对任意的 $\bar{x} \in \Omega - F_\infty$, $\mathcal{T}_\Omega^F$ 在 $\bar{x}$ 处的次微分可通过

$$\partial\mathcal{T}_\Omega^F(\bar{x}) = N(\bar{x}; \Omega - F_\infty) \cap C^* \tag{3.16}$$

计算, 这里 $C^* := \{v \in \mathbb{R}^n | \sigma_F(-v) \leqslant 1\}$ 通过 (2.41) 定义. 特别地, 有

$$\partial\mathcal{T}_\Omega^F(\bar{x}) = N(\bar{x}; \Omega) \cap C^*, \quad \forall \bar{x} \in \Omega. \tag{3.17}$$

证明　任取 $v \in \partial\mathcal{T}_\Omega^F(\bar{x})$. 利用式 (2.13) 和事实 $\mathcal{T}_\Omega^F(x) = 0, \forall x \in \Omega - F_\infty$, 得 $v \in N(\bar{x}; \Omega - F_\infty)$. 对 $\bar{x} \in \Omega - F_\infty$, $\bar{x} = \bar{\omega} - d$, 其中 $\bar{\omega} \in \Omega, d \in F_\infty$. 任取 $f \in F$, 并定义 $x_t := (\bar{x} + d) - tf = \bar{\omega} - tf$, $t > 0$. 于是 $(x_t + tF) \cap \Omega \neq \varnothing$, 因此

$$\langle v, x_t - \bar{x} \rangle \leqslant \mathcal{T}_\Omega^F(x_t) \leqslant t, \quad \text{或等价地}, \quad \left\langle v, \frac{d}{t} - f \right\rangle \leqslant 1, \quad \forall t > 0.$$

通过令 $t \to \infty$, 这蕴涵着 $\langle v, -f \rangle \leqslant 1$, 因此得 $v \in C^*$. 从而包含关系 $\partial\mathcal{T}_\Omega^F(\bar{x}) \subset N(\bar{x}; \Omega - F_\infty) \cap C^*$ 成立.

为证式 (3.16) 中反向的包含关系成立, 任取 $v \in N(\bar{x}; \Omega - F_\infty) \cap C^*$, $u \in \operatorname{dom} \mathcal{T}_\Omega^F$. 对任意的 $\varepsilon > 0$, 存在 $t \in [0, \infty), \omega \in \Omega$ 和 $f \in F$ 使得

$$\mathcal{T}_\Omega^F(u) \leqslant t < \mathcal{T}_\Omega^F(u) + \varepsilon, \quad \text{且} \quad u + tf = \omega.$$

由于 $\Omega \subset \Omega - F_\infty$, 有关系式

$$\langle v, u - \bar{x} \rangle = \langle v, \omega - \bar{x} \rangle + t\langle v, -f \rangle \leqslant t < \mathcal{T}_\Omega^F(u) + \varepsilon = \mathcal{T}_\Omega^F(u) - \mathcal{T}_\Omega^F(\bar{x}) + \varepsilon,$$

这表明 $v \in \mathcal{T}_\Omega^F(\bar{x})$. 这样可得

$$N(\bar{x}; \Omega - F_\infty) \cap C^* \subset \partial\mathcal{T}_\Omega^F(\bar{x}),$$

从而证明了一般情形下, 当 $\bar{x} \in \Omega - F_\infty$ 时次微分公式 (3.16) 成立.

余下证明当 $\bar{x} \in \Omega$ 时简化的表达式 (3.17) 成立. 由于 $\Omega \subset \Omega - F_\infty$, 显然有包含关系

$$N(\bar{x}; \Omega - F_\infty) \cap C^* \subset N(\bar{x}; \Omega) \cap C^*.$$

为证反向的包含关系成立, 取 $v \in N(\bar{x};\Omega) \cap C^*$, 取定 $x \in \Omega - F_\infty$. 取 $\omega \in \Omega, d \in F_\infty$ 使得 $x = \omega - d$, 由于 $0 \in F$, 所以 $td \in F$, 因此 $\langle v, -td\rangle \leqslant 1, \forall t > 0$. 于是 $\langle v, -d\rangle \leqslant 0$, 从而蕴涵着

$$\langle v, x - \bar{x}\rangle = \langle v, \omega - d - \bar{x}\rangle = \langle v, \omega - \bar{x}\rangle + \langle v, -d\rangle \leqslant 0.$$

这样可得 $v \in N(\bar{x};\Omega - F_\infty) \cap C^*$, 从而完成了证明. □

下面的定理给出与定理 3.37 互补的情形, 即当 $\bar{x} \notin \Omega - F_\infty$ 且 $\bar{x} \notin \Omega$ 时极小时间函数的次微分计算.

定理 3.38 设集合 $0 \in F \subset \mathbb{R}^n$ 和 $\Omega \subset \mathbb{R}^n$ 都是非空闭凸的, 设 $\bar{x} \in \operatorname{dom} \mathcal{T}_\Omega^F$ 为极小时间函数 (3.9) 的定义域. 假设 Ω 是有界的, 则对 $\bar{x} \notin \Omega - F_\infty$, 有

$$\partial \mathcal{T}_\Omega^F(\bar{x}) = N(\bar{x};\Omega_r) \cap \xi^*. \tag{3.18}$$

这里扩张 Ω_r 由式 (3.15) 定义, 而

$$S^* := \{v \in \mathbb{R}^n | \sigma_F(-v) = 1\}, \quad r = \mathcal{T}_\Omega^F(\bar{x}) > 0.$$

证明 取 $v \in \partial \mathcal{T}_\Omega^F(\bar{x})$. 由定义得

$$\langle v, x - \bar{x}\rangle \leqslant \mathcal{T}_\Omega^F(x) - \mathcal{T}_\Omega^F(\bar{x}), \quad \forall x \in \mathbb{R}^n. \tag{3.19}$$

于是 $\mathcal{T}_\Omega^F(x) \leqslant r = \mathcal{T}_\Omega^F(\bar{x})$, 因此 $\langle v, x - \bar{x}\rangle \leqslant 0, \forall x \in \Omega_r$. 从而 $v \in N(\bar{x};\Omega_r)$. 在式 (3.19) 中令 $x := \bar{x} - f$, 其中 $f \in F$, 则命题 3.36 保证 $\sigma_F(-v) \leqslant 1$. 取定 $\varepsilon \in (0, r)$, 选取 $t \in \mathbb{R}, f \in F$ 和 $\omega \in \Omega$ 使得

$$r \leqslant t < r + \varepsilon^2 \quad \text{和} \quad \omega = \bar{x} + tf.$$

因为 $\omega = \bar{x} + \varepsilon f + (t - \varepsilon)f$, 则得 $\mathcal{T}_\Omega^F(\bar{x} + \varepsilon f) \leqslant t - \varepsilon$. 在式 (3.19) 中令 $x := \bar{x} + \varepsilon f$, 得估计

$$\langle v, \varepsilon f\rangle \leqslant \mathcal{T}_\Omega^F(\bar{x} + \varepsilon f) - \mathcal{T}_\Omega^F(\bar{x}) \leqslant t - \varepsilon - r \leqslant \varepsilon^2 - \varepsilon.$$

因此有

$$1 - \varepsilon \leqslant \langle -v, f\rangle \leqslant \sigma_F(-v).$$

现令 $\varepsilon \downarrow 0$ 得 $\sigma_F(-v) \geqslant 1$, 即此时 $\sigma_F(-v) = 1$. 因此 $v \in \xi^*$, 这就证明了包含关系 $\partial \mathcal{T}_\Omega^F(\bar{x}) \subset N(\bar{x};\Omega_r) \cap \xi^*$ 成立.

为证式 (3.18) 中反向的包含关系成立, 任取 $v \in N(\bar{x};\Omega_r)$ 且 $\sigma_F(-v) = 1$, 下面证明式 (3.19) 成立. 由定理 3.37 得 $v \in \partial \mathcal{T}_{\Omega_r}^F(\bar{x})$, 因此

$$\langle v, x - \bar{x}\rangle \leqslant \mathcal{T}_{\Omega_r}^F(x), \quad \forall x \in \mathbb{R}^n.$$

取定 $x \in \mathbb{R}^n$, 由命题 3.35 推得当 $t := \mathcal{T}_\Omega^F(x) > r$ 时 $\mathcal{T}_\Omega^F(x) - r = \mathcal{T}_{\Omega_r}^F(x)$, 这就保证式 (3.19) 成立. 现假设 $t \leqslant r$, 对任意的 $\varepsilon > 0$, 选取 $f \in F$ 使得 $\langle v, -f\rangle > 1 - \varepsilon$, 则由命题 3.36 得 $\mathcal{T}_\Omega^F(x - (r-t)f) \leqslant r$, 因此 $x - (r-t)f \in \Omega_r$. 由于 $v \in N(\bar{x}; \Omega_r)$, 有 $\langle v, x - (r-t)f - \bar{x}\rangle \leqslant 0$, 从而有

$$\langle v, x - \bar{x}\rangle \leqslant \langle v, f\rangle(r-t) \leqslant (1-\varepsilon)(t-r) = (1-\varepsilon)(\mathcal{T}_\Omega^F(x) - \mathcal{T}_\Omega^F(\bar{x})).$$

由 $\varepsilon > 0$ 的任意性, 得式 (3.19) 成立, 因此证明了 $v \in \partial \mathcal{T}_\Omega^F(\bar{x})$. □

下面的结果处理与定理 3.38 情形相同, 但利用 Minkowski 度规的次微分和到目标集合的广义投影给出计算极小时间函数次微分的另一公式.

定理 3.39　设集合 $0 \in F \subset \mathbb{R}^n$ 和 $\Omega \subset \mathbb{R}^n$ 都是非空闭凸的, 设 $\bar{x} \in \operatorname{dom} \mathcal{T}_\Omega^F$. 假设 Ω 是有界的且 $\bar{x} \notin \Omega - F_\infty$, 则有

$$\partial \mathcal{T}_\Omega^F(\bar{x}) = [-\partial \rho_F(\bar{\omega} - \bar{x})] \cap N(\bar{\omega}; \Omega) \tag{3.20}$$

对任意的 $\bar{\omega} \in \Pi_F(\bar{x}; \Omega)$ 成立, 其中的广义投影定义为

$$\Pi_F(\bar{x}; \Omega) := \{\omega \in \Omega \mid \mathcal{T}_\Omega^F(\bar{x}) = \rho_F(\omega - \bar{x})\}. \tag{3.21}$$

证明　任取 $v \in \partial \mathcal{T}_\Omega^F(\bar{x}), \bar{\omega} \in \Pi_F(\bar{x}; \Omega)$, 则 $\mathcal{T}_\Omega^F(\bar{x}) = \rho_F(\bar{\omega} - \bar{x})$ 且

$$\langle v, x - \bar{x}\rangle \leqslant \mathcal{T}_\Omega^F(x) - \mathcal{T}_\Omega^F(\bar{x}), \quad \forall\, x \in \mathbb{R}^n. \tag{3.22}$$

易验证下面的关系式对任意的 $\omega \in \Omega$ 成立:

$$\begin{aligned}\langle v, \omega - \bar{\omega}\rangle &= \langle v, (\omega - \bar{\omega} + \bar{x}) - \bar{x}\rangle \leqslant \mathcal{T}_\Omega^F(\omega - \bar{\omega} + \bar{x}) - \mathcal{T}_\Omega^F(\bar{x}) \\ &\leqslant \rho_F(\omega - (\omega - \bar{\omega} + \bar{x})) - \rho_F(\bar{\omega} - \bar{x}) = 0.\end{aligned}$$

这表明 $v \in N(\bar{\omega}; \Omega)$. 记 $\tilde{u} := \bar{\omega} - \bar{x}$, 对任意的 $t \in (0,1), u \in \mathbb{R}^n$, 在式 (3.22) 中令 $x := \bar{x} - t(u - \tilde{u})$, 有

$$\begin{aligned}\langle v, -t(u - \tilde{u})\rangle &\leqslant \mathcal{T}_\Omega^F(\bar{x} - t(u - \tilde{u})) - \mathcal{T}_\Omega^F(\bar{x}) \\ &\leqslant \rho_F(\bar{\omega} - (\bar{x} - t(u - \tilde{u}))) - \rho_F(\bar{\omega} - \bar{x}) \\ &= \rho_F(\tilde{u} + t(u - \tilde{u})) - \rho_F(\tilde{u}) \\ &= \rho_F(tu + (1-t)\tilde{u}) - \rho_F(\tilde{u}) \\ &\leqslant t\rho_F(u) + (1-t)\rho_F(\tilde{u}) - \rho_F(\tilde{u}) \\ &= t(\rho_F(u) - \rho_F(\tilde{u})).\end{aligned}$$

因此

$$\langle -v, u - \tilde{u}\rangle \leqslant \rho_F(u) - \rho_F(\tilde{u}), \quad \forall\, u \in \mathbb{R}^n.$$

于是 $-v \in \partial\rho_F(\tilde{u})$, 因此 $v \in -\partial\rho_F(\bar{\omega}-\bar{x})$, 这就证明了式 (3.20) 中的包含关系"$\subset$".

为证反向的包含关系成立, 记

$$\rho_F(x) = \inf\{t \geqslant 0 \mid x \in tF\} = \inf\{t \geqslant 0 \mid (-x+tF) \cap O \neq \varnothing\} = \mathcal{T}_O^F(-x).$$

其中 $O := \{0\}$. 由于 $\bar{x} \notin \Omega - F_\infty$, 对任意的 $\bar{\omega} \in \Omega$, 有 $\bar{x}-\bar{\omega} \notin O - F_\infty = -F_\infty$. 进一步, 利用定理 3.38 得 $-v \in \partial\rho_F(\bar{\omega}-\bar{x})$ 从而得 $v \in \xi^*$. 因此余下需证包含关系

$$[-\partial\rho_F(\bar{\omega}-\bar{x})] \cap N(\bar{\omega};\Omega) \subset N(\bar{x};\Omega_r) \tag{3.23}$$

成立, 再应用定理 3.38. 为证式 (3.23) 成立, 任取向量 $x \in \Omega_r$, 任意小的正数 ε, 则存在 $t < r+\varepsilon, f \in F$ 和 $\omega \in \Omega$ 使得 $\omega = x+tf$, 因此有

$$\begin{aligned}
\langle v, x-\bar{x}\rangle &= \langle v, \omega - tf - \bar{x}\rangle \\
&= t\langle -v, f\rangle + \langle v, \omega-\bar{\omega}\rangle + \langle v, \bar{\omega}-\bar{x}\rangle \\
&\leqslant t + \langle v, \omega-\bar{\omega}\rangle + \langle v, \bar{\omega}-\bar{x}\rangle \\
&\leqslant \mathcal{T}_\Omega^F(\bar{x}) + \varepsilon + \langle v, \omega-\bar{\omega}\rangle + \langle v, \bar{\omega}-\bar{x}\rangle.
\end{aligned}$$

由于 $v \in N(\bar{\omega};\Omega)$, 注意到 $\langle v, \omega-\bar{\omega}\rangle \leqslant 0$. 而且, 由于 $-v \in \partial\rho_F(\bar{\omega}-\bar{x})$, 所以

$$\langle v, \bar{\omega}-\bar{x}\rangle = \langle -v, 0-(\bar{\omega}-\bar{x})\rangle \leqslant \rho_F(0) - \rho_F(\bar{\omega}-\bar{x}) = -\mathcal{T}_\Omega^F(\bar{x}).$$

因此对任意的 $x \in \Omega_r, \langle v, x-\bar{x}\rangle \leqslant \varepsilon$. 由 $\varepsilon > 0$ 的任意性, 得 $v \in N(\bar{\omega};\Omega_r)$. □

3.9 Nash 均 衡

本节给出非合作对策理论的简要引论, 并给出 Nash 均衡存在性的一个简单证明, 作为凸性和 Brouwer 不动点定理的推论. 为简单起见, 我们只考虑二人对策, 而更一般情形可类似地处理.

定义 3.40 设 Ω_1 和 Ω_2 分别是 $\mathbb{R}^n$ 和 $\mathbb{R}^p$ 的非空子集, 针对两个选手 I 和 II 的非合作对策由两个策略集 Ω_i, 两个实值函数 $u_i : \Omega_1 \times \Omega_2 \to \mathbb{R}\,(i=1,2)$ 组成, u_i 称为支付函数. 把这个对策记为 $\{\Omega_i, u_i\}$, $i=1,2$.

Nash 均衡这个关键概念由 Jr John Forbes Nash 于 1950 年引入, 他证明了这个均衡的存在性并且因此于 1994 年获得了 Nobel 经济学奖.

定义 3.41 给定非合作二人对策 $\{\Omega_i, u_i\}, i=1,2$. Nash 均衡是指元素 $(\bar{x}_1, \bar{x}_2) \in \Omega_1 \times \Omega_2$, 满足条件

$$\begin{aligned}
u_1(x_1, \bar{x}_2) &\leqslant u_1(\bar{x}_1, \bar{x}_2), \quad \forall\, x_1 \in \Omega_1, \\
u_2(\bar{x}_1, x_2) &\leqslant u_2(\bar{x}_1, \bar{x}_2), \quad \forall\, x_2 \in \Omega_2.
\end{aligned}$$

上面对最优策略对 $(\bar{x}_1, \bar{x}_2)$ 给出的条件的意思是, 使得当选手 II 选择策略 $\bar{x}_2$ 时 $\bar{x}_1$ 是选手 I 的最好回应, 以及当选手 I 选择策略 $\bar{x}_1$ 时 $\bar{x}_2$ 是选手 II 的最好回应, 在此意义下两个选手达成一致的最优策略. 下面考虑两个例子来解释说明这个概念.

例 3.42　在二人对策中假设每个选手只能选择策略 A 或策略 B. 如果两个选手都选择策略 A, 他们都得 4 个点. 如果选手 I 选择策略 A 而选手 II 选择策略 B, 那么选手 I 得 1 个点而选手 II 得 3 个点. 如果选手 I 选择策略 B 而选手 II 选择策略 A, 那么选手 I 得 3 个点而选手 II 得 1 个点. 如果两个选手都选策略 B, 每个选手都得 2 个点. 每个选手的支付函数用下面的支付矩阵表示为

		选手 II	
		A	B
选手 I	A	4,4	1,3
	B	3,1	2,2

在例 3.42 中, 当两个选手都选择策略 A 时 Nash 均衡出现, 而且当两个选手都选择策略 B 时 Nash 均衡也出现. 当我们考虑, 例如, 当两个选手都选择策略 B 的情形. 此时, 选手 II 选择策略 B 已定, 选手 I 也想保住策略 B, 因为策略的改变将导致他/她的支付从 2 减少到 1. 同理, 选手 I 选择策略 B 已定, 选手 II 想保住策略 B, 因为策略的改变也将导致他/她的支付从 2 减少到 1.

例 3.43　让我们考虑另一个简单的二人对策, 称为便士匹配 (matching pennies). 假设选手 I 和选手 II 每个人都有一个便士. 每个选手必须偷偷地将便士翻为正面或反面, 然后同时显示他/她的选择. 如果两个硬币显示相同的面 (正－正, 或反－反), 那么选手 I 赢得比赛并得到选手 II 的便士. 在硬币显示不同的面 (正－反或反－正) 的另一种情形下, 选手 II 赢得比赛并得到选手 I 的便士. 每个选手的支付函数通过下面的支付矩阵表示为

		选手 II	
		正	反
选手 I	正	1,−1	−1,1
	反	−1,1	1,−1

在这个比赛中, 不难看到没有 Nash 均衡存在.

用 A 矩阵表示选手 I 的支付函数, 并用 B 矩阵表示选手 II 的支付函数:

$$A = \begin{array}{|c|c|}\hline 1 & -1 \\ \hline -1 & 1 \\ \hline \end{array}, \quad B = \begin{array}{|c|c|}\hline -1 & 1 \\ \hline 1 & -1 \\ \hline \end{array},$$

现假设选手 I 正与一个知道选手 I 所选择的面的知心术者比赛. 如果选手 I 决定翻为正面, 那么选手 II 知道这种情况并翻为反面从而赢得比赛. 当选手 I 决定翻为反面的情形, 则选手 II 选择翻为正面从而再次赢得比赛. 因此为保证比赛的公平, 他/她决定把他们策略随机化. 例如, 改成抛硬币而不是翻硬币. 描述这个新的比赛如下

		选手 II	
		正, q_1	反, q_2
选手 I	正, p_1	1,−1	−1,1
	反, p_2	−1,1	1,−1

在这个新的比赛中, 选手 I 的硬币出现正面的概率为 p_1, 出现反面的概率为 p_2, 其中 $p_1+p_2=1$. 类似地, 选手 II 随便地利用另一硬币且出现头部的概率为 q_1, 出现反面的概率为 q_2, 其中 $q_1+q_2=1$. 这个新的策略称为混合策略而最初的那个策略称为纯策略.

假设选手 II 应用混合策略 $\{q_1,q_2\}$, 则选手 I 选取正面这个纯策略的期望支付是

$$u_H(q_1,q_2)=q_1-q_2=2q_1-1.$$

类似地, 选手 I 选取反面这个纯策略的期望支付是

$$u_T(q_1,q_2)=-q_1+q_2=-2q_1+1.$$

因此选手 I 关于混合策略 $\{p_1,p_2\}$ 的期望支付是

$$u_1(p,q)=p_1u_H(q_1,q_2)+p_2u_T(q_1,q_2)=p_1(q_1-q_2)+p_2(-q_1+q_2)=p^{\mathrm T}Aq,$$

其中 $p=[p_1,p_2]^{\mathrm T},q=[q_1,q_2]^{\mathrm T}$.

同理, 如果选手 I 选择混合策略 $\{p_1,p_2\}$, 那么选手 II 关于混合策略 $\{q_1,q_2\}$ 的期望支付是

$$u_2(p,q)=p^{\mathrm T}Bq=-u_1(p,q).$$

对这个新的比赛来说, 元素 $(\bar p,\bar q)\in\Delta\times\Delta$ 是一个 Nash 均衡, 如果

$$u_1(p,\bar q)\leqslant u_1(\bar p,\bar q),\quad \forall p\in\Delta,$$

$$u_2(\bar p,q)\leqslant u_2(\bar p,\bar q),\quad \forall q\in\Delta,$$

这里 $\Delta:=\{(p_1,p_2)|p_1\geqslant 0,p_2\geqslant 0,p_1+p_2=1\}$ 是一个非空紧凸集.

Nash[23,24] 在更一般的情形下证明了关于混合策略的均衡的存在性. 他的证明基于 Brouwer 不动点定理, 且过程相当复杂. 现在沿用 Rockafellar[28] 的方法, 给出

Nash 均衡定理的一个更加简单的证明, 也利用 Brouwer 不动点定理, 但其他方面仅仅利用凸分析和最优化的基本工具.

定理 3.44　考虑二人对策 $\{\Omega_i, u_i\}$, 其中 $\Omega_1 \subset \mathbb{R}^n, \Omega_2 \subset \mathbb{R}^p$ 是非空紧凸集. 设支付函数 $u_i : \Omega_1 \times \Omega_2 \to \mathbb{R}$ 由

$$u_1(p, q) := p^{\mathrm{T}} A q, \quad u_2(p, q) := p^{\mathrm{T}} B q$$

给出, 这里 A 和 B 是 $n \times p$ 矩阵, 则这个对策有一个 Nash 均衡.

证明　由定义得元素 $(\bar{p}, \bar{q}) \in \Omega_1 \times \Omega_2$ 是所考虑对策的一个 Nash 均衡, 当且仅当

$$-u_1(p, \bar{q}) \geqslant -u_1(\bar{p}, \bar{q}), \quad \forall\, p \in \Omega_1,$$

$$-u_2(\bar{p}, q) \geqslant -u_2(\bar{p}, \bar{q}), \quad \forall\, q \in \Omega_2.$$

作为一个简单的练习可以证明 (也可参见推论 4.15) 这个结论成立当且仅当有法锥包含关系

$$\nabla_p u_1(\bar{p}, \bar{q}) \in N(\bar{p}; \Omega_1), \quad \nabla_q u_2(\bar{p}, \bar{q}) \in N(\bar{q}; \Omega_2). \tag{3.24}$$

通过利用 u_i 的结构并定义 $\Omega := \Omega_1 \times \Omega_2$, 条件式 (3.24) 可用下面等价的方法之一表达:

$$A\bar{q} \in N(\bar{p}; \Omega_1), \quad 且 B^{\mathrm{T}} \bar{p} \in N(\bar{q}; \Omega_2);$$

$$(A\bar{q}, B^{\mathrm{T}} \bar{p}) \in N(\bar{p}; \Omega_1) \times N(\bar{q}; \Omega_2) = N((\bar{p}, \bar{q}); \Omega_2 \times \Omega_2) = N((\bar{p}, \bar{q}); \Omega).$$

利用关于 Euclid 投影的命题 1.78, 这也等价于

$$(\bar{p}, \bar{q}) = \Pi((\bar{p}, \bar{q}) + (A\bar{q}, B^{\mathrm{T}} \bar{p}); \Omega). \tag{3.25}$$

现定义映射 $F : \Omega \to \Omega$ 为

$$F(p, q) := \Pi((p, q) + (Aq, B^{\mathrm{T}} p); \Omega), \quad 其中 \Omega := \Omega_1 \times \Omega_2,$$

并利用命题 1.79 中另一基本的投影性质可得映射 F 是连续的, 而且集合 Ω 是非空紧凸的. 利用经典的 Brouwer 不动点定理 (例如, 参见文献 [33]), 映射 F 有不动点 $(\bar{p}, \bar{q}) \in \Omega$, 即式 (3.25) 成立, 因此它是该对策的一个 Nash 均衡. □

3.10　练　　习

练习 3.1　设函数 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 在 $\bar{x}$ 附近有限. 证明 f 在 $\bar{x}$ 处是 Gâteaux 可微的, 且 $f'_G(\bar{x}) = v$ 当且仅当方向导数 $f'(\bar{x}; d)$ 存在且 $f'(\bar{x}; d) = \langle v, d \rangle, \forall\, d \in \mathbb{R}^n$.

练习 3.2　设 Ω 是 $\mathbb{R}^n$ 的非空闭凸子集.

(i) 证明函数 $f(x) := [d(x;\Omega)]^2$ 在 $\mathbb{R}^n$ 上是可微的, 且

$$\nabla f(\bar{x}) = 2[\bar{x} - \Pi(\bar{x};\Omega)]$$

对任意的 $\bar{x} \in \mathbb{R}^n$ 成立.

(ii) 证明函数 $g(x) := d(x;\Omega)$ 在 $\mathbb{R}^n \backslash \Omega$ 上是可微的.

练习 3.3　设 F 是 $\mathbb{R}^n$ 的非空紧凸子集, 设 $A : \mathbb{R}^n \to \mathbb{R}^p$ 是线性映射. 定义

$$f(x) := \sup_{u \in F} \left(\langle Ax, u \rangle - \frac{1}{2}\|u\|^2 \right),$$

证明 f 是凸的和可微的.

练习 3.4　构造一个函数 $f : \mathbb{R}^n \to \mathbb{R}$ 的例子, 它是 Gâteaux 可微的, 但不是 Fréchet 可微的. 这样的函数能是凸的吗?

练习 3.5　设 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸函数, 设 $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$. 证明: 如果 $\dfrac{\partial f}{\partial x_i}$ 在 $\bar{x}$ 处存在, $i = 1, \cdots, n$, 则 f 在 $\bar{x}$ 处是 (Fréchet) 可微的.

练习 3.6　设 $f : \mathbb{R}^n \to \mathbb{R}$ 在 $\mathbb{R}^n$ 上是可微的, 证明 f 是严格凸的当且仅当它满足条件

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle < f(x) - f(\bar{x}), \quad \forall\, x \in \mathbb{R}^n, \bar{x} \in \mathbb{R}^n, x \neq \bar{x}.$$

练习 3.7　设 $f : \mathbb{R}^n \to \mathbb{R}$ 在 $\mathbb{R}^n$ 上是二次连续可微 ($\mathcal{C}^2$) 的. 证明: 如果对任意的 $\bar{x} \in \mathbb{R}^n$, Hesse 矩阵 $\nabla^2 f(\bar{x})$ 是正定的, 则 f 是严格凸的. 给一个例子说明反向的结论一般来说不成立.

练习 3.8　凸函数 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 称为强凸的且具有模 $c > 0$, 如果对任意的 $x_1, x_2 \in \mathbb{R}^n, \lambda \in (0,1)$, 有

$$f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f(x_1) + (1-\lambda) f(x_2) - \frac{1}{2} c\lambda(1-\lambda)\|x_1 - x_2\|^2.$$

证明: 对任意的 $\mathcal{C}^2$, 凸函数 $f : \mathbb{R}^n \to \mathbb{R}$, 下列条件是等价的:

(i) f 是强凸的.

(ii) $f - \dfrac{c}{2}\|\cdot\|^2$ 是凸的.

(iii) $\langle \nabla^2 f(x) d, d \rangle \geqslant c\|d\|^2, \forall\, x, d \in \mathbb{R}^n$.

练习 3.9　集值映射 $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ 称为单调的, 如果

$$\langle v_2 - v_1, x_2 - x_1 \rangle \geqslant 0, \quad \forall\, v_i \in F(x_i), \quad i = 1, 2.$$

F 称为严格单调的, 如果

$$\langle v_2 - v_1, x_2 - x_1 \rangle > 0, \quad \forall\, v_i \in F(x_i), x_1 \neq x_2, i = 1, 2.$$

它称为具有模 $\ell>0$ 强单调的, 如果

$$\langle v_2-v_1,x_2-x_1\rangle\geqslant \ell\|x_2-x_1\|^2,\quad \forall v_i\in F(x_i),i=1,2.$$

(i) 证明: 对任何凸函数 $f:\mathbb{R}^n\to\mathbb{R}$, 映射 $F(x):=\partial f(x)$ 是单调的.

(ii) 进一步, 对任何凸函数 $f:\mathbb{R}^n\to\mathbb{R}$, 次微分映射 $\partial f(x)$ 是严格单调的当且仅当 f 是严格凸的, 并且它是强单调的当且仅当 f 是强凸的.

练习 3.10　证明推论 3.7.

练习 3.11　设 A 是一个 $p\times n$ 矩阵, 设 $b\in\mathbb{R}^p$. 应用 Farkas 引理来证明系统 $Ax=b,x\geqslant 0$ 有可行解 $x\in\mathbb{R}^n$ 当且仅当系统 $A^{\mathrm{T}}y\leqslant 0,b^{\mathrm{T}}y>0$ 无可行解 $y\in\mathbb{R}^p$.

练习 3.12　从 Radon 定理及其证明推出 Carathéodory 定理.

练习 3.13　利用 Carathéodory 定理来证明 Helly 定理.

练习 3.14　设 C 是 $\mathbb{R}^n$ 的子集. 证明 C 的极化可表示为

$$C^\circ=C^\perp:=\{v\in\mathbb{R}^n|\langle v,c\rangle=0,\forall\, c\in C\}.$$

练习 3.15　设 C 是 $\mathbb{R}^n$ 的非空凸锥. 证明

$$C^\circ=\{v\in\mathbb{R}^n|\langle v,c\rangle\leqslant 1,\forall\, c\in C\}.$$

练习 3.16　设 Ω_1 和 Ω_2 是凸集且 $\operatorname{int}\Omega_1\cap\Omega_2\neq\varnothing$, 证明

$$T(\bar{x};\Omega_1\cap\Omega_2)=T(\bar{x};\Omega_1)\cap T(\bar{x};\Omega_2),\quad \forall\bar{x}\in\Omega_1\cap\Omega_2.$$

练习 3.17　设 A 是 $p\times n$ 矩阵, 设

$$\Omega:=\{x\in\mathbb{R}^n|Ax=0,x\geqslant 0\},$$

而 $\bar{x}=(\bar{x}_1,\cdots,\bar{x}_n)\in\Omega$. 证明切锥和法锥公式:

$$T(\bar{x};\Omega)=\{u=(u_1,\cdots,u_n)\in\mathbb{R}^n|Au=0,\text{若}\bar{x}_i=0,\text{则}u_i\geqslant 0\},$$
$$N(\bar{x};\Omega)=\{v\in\mathbb{R}^n|v=A^*y-z,y\in\mathbb{R}^p,z\in\mathbb{R}^n,z_i\geqslant 0,\langle z,\bar{x}\rangle=0\}.$$

练习 3.18　给一个闭凸集 $\Omega\subset\mathbb{R}^2$ 的例子, 使得在某点 $\bar{x}\in\Omega$ 使得锥 $\mathbb{R}^+(\Omega-\bar{x})$ 不是闭的.

练习 3.19　设 $f:\mathbb{R}^n\to\mathbb{R}$ 是凸函数且存在 $\ell\geqslant 0$ 使得 $\partial f(x)\subset\mathbb{B}$. 证明 f 在 $\mathbb{R}^n$ 上是 ℓ-Lipschitz 连续的.

练习 3.20　设 $f:\mathbb{R}\to\mathbb{R}$ 是凸的. 证明 f 是非减的当且仅当 $\partial f(x)\subset[0,\infty),\forall\, x\in\mathbb{R}$.

练习 3.21 求下列集合的地平锥:

(i) $F := \{(x,y) \in \mathbb{R}^2 | y \geqslant x^2\}$.

(ii) $F := \{(x,y) \in \mathbb{R}^2 | y \geqslant |x|\}$.

(iii) $F := \left\{(x,y) \in \mathbb{R}^2 \middle| x > 0, y \geqslant \dfrac{1}{x}\right\}$.

(iv) $F := \{(x,y) \in \mathbb{R}^2 | x > 0, y \geqslant \sqrt{x^2+1}\}$.

练习 3.22 设 $v_i \in \mathbb{R}^n$, $b_i > 0$, $i = 1, \cdots, m$, 考虑集合

$$F := \{x \in \mathbb{R}^n | \langle v_i, x\rangle \leqslant b_i, i = 1, \cdots, m\}. \tag{3.26}$$

证明 F 的地平锥和 Minkowski 度规的如下表示:

$$F_\infty := \{x \in \mathbb{R}^n | \langle v_i, x\rangle \leqslant 0, i = 1, \cdots, m\},$$

$$\rho_F(u) = \max\left\{0, \max_{i\in\{1,\cdots,m\}} \frac{\langle v_i, u\rangle}{b_i}\right\}.$$

练习 3.23 设 A, B 是 $\mathbb{R}^n$ 的非空闭凸子集.

(i) 证明 $(A \times B)_\infty = A_\infty \times B_\infty$.

(ii) 求条件保证 $(A+B)_\infty = A_\infty + B_\infty$.

练习 3.24 设 $F \neq \varnothing$ 是闭有界凸集, 且设 $\Omega \neq \varnothing$ 是 $\mathbb{R}^n$ 的闭集. 证明 $\mathcal{T}_\Omega^F(x) = 0$ 当且仅当 $x \in \Omega$. 这意味着定理 3.27(i) 中的条件 $0 \in F$ 在这种情况下不是必要的.

练习 3.25 设 $F \neq \varnothing$ 是闭有界凸集, 设 $\Omega \neq \varnothing$ 是闭凸集. 分别考虑定理 3.37 和定理 3.38 中定义的集合 C^* 和 S^*. 证明下列结论成立:

(i) 如果 $\bar{x} \in \Omega$, 那么

$$\partial\mathcal{T}_\Omega^F(\bar{x}) = N(\bar{x}; \Omega) \cap C^*.$$

(ii) 如果 $\bar{x} \notin \Omega$, 那么

$$\partial\mathcal{T}_\Omega^F(\bar{x}) = N(\bar{x}; \Omega_r) \cap S^*, \quad r := \mathcal{T}_\Omega^F(\bar{x}) > 0.$$

(iii) 如果 $\bar{x} \notin \Omega$, 那么

$$\partial\mathcal{T}_\Omega^F(\bar{x}) = [-\partial\rho_F(\bar{\omega} - \bar{x})] \cap N(\bar{\omega}; \Omega)$$

对任意的 $\bar{\omega} \in \Pi_F(\bar{x}; \Omega)$ 成立.

练习 3.26 给出 Nash 均衡的刻画式 (3.24) 的一个直接证明.

练习 3.27 在例 3.43 混合策略中, 找到给出这个对策的一个 Nash 均衡的 $(\bar{p}, \bar{q})$.

第 4 章　在最优化和选址问题中的应用

本章给出第 1~3 章中凸分析结果在凸最优化和选址问题一些选题中的应用. 在最优化问题中的应用主要是经典的, 但是以简化和统一的方式给出. 另外, 在选址问题中的应用主要基于最近期刊发表的论文, 在专著和教材中还没有从凸分析的角度系统地讨论.

4.1　下半连续性和极小值点的存在性

首先讨论增广实值函数下半连续性概念, 它在证明绝对极小值点的存在性和相关课题中起着至关重要的作用. 值得一提的是, 下面给出的许多结果不要求 f 的凸性.

定义 4.1　称增广实值函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 在 $\bar{x}\in\mathbb{R}^n$ 处是下半连续的 (l.s.c.), 如果对任意的满足 $f(\bar{x})>\alpha$ 的 $\alpha\in\mathbb{R}$, 存在 $\delta>0$ 使得

$$f(x)>\alpha,\quad \forall\, x\in\mathbb{B}(\bar{x};\delta). \tag{4.1}$$

简称 f 是下半连续的, 如果它在 $\mathbb{R}^n$ 的每一点处都是 l.s.c..

下面给出一些下半连续性成立或不成立的例子.

例 4.2　(i) 定义函数 $f:\mathbb{R}\to\mathbb{R}$ 为

$$f(x):=\begin{cases} x^2, & x\neq 0,\\ -1, & x=0.\end{cases}$$

易见它在 $\bar{x}=0$ 处是 l.s.c.的, 实际上它在整个实直线上是 l.s.c. 的. 另外, 用 $f(0)=1$ 替代 $f(0)=-1$, 则破坏了在 $\bar{x}=0$ 处的下半连续性 (图 4.1).

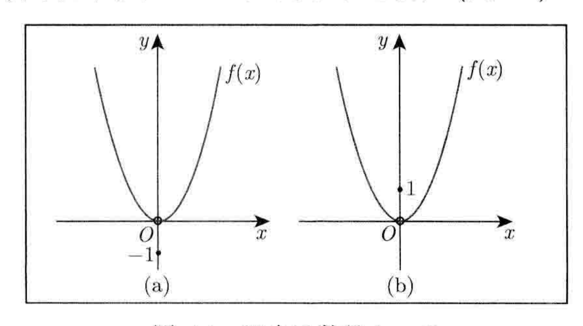


图 4.1　哪个函数是 l.s.c.?

(ii) 考虑集合 $[-1,1]\subset\mathbb{R}$ 的指示函数:

$$f(x):=\begin{cases}0, & |x|\leqslant 1,\\ \infty, & \text{其他}\end{cases}$$

给出, 则 f 在 $\mathbb{R}$ 上是下半连续函数. 然而, 小的改动

$$g(x):=\begin{cases}0, & |x|< 1,\\ \infty, & \text{其他},\end{cases}$$

则破坏了在 $\mathbb{R}^n$ 上的下半连续性. 事实上, g 在 $x=\pm 1$ 处不是 l.s.c., 而在实直线的任何其他点处它是下半连续的.

下面四个命题给出下半连续性有用的刻画.

命题 4.3 设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$, $\bar{x}\in\operatorname{dom} f$, 则 f 在 $\bar{x}$ 处是 l.s.c.的当且仅当对任意的 $\varepsilon>0$, 存在 $\delta>0$ 使得

$$f(\bar{x})-\varepsilon<f(x),\quad \forall\, x\in\mathbb{B}(\bar{x};\delta). \tag{4.2}$$

证明 假设 f 在 $\bar{x}$ 是 l.s.c. 的. 由于 $\lambda:=f(\bar{x})-\varepsilon<f(\bar{x})$, 存在 $\delta>0$, 使得

$$f(\bar{x})-\varepsilon=\lambda<f(x),\quad \forall\, x\in\mathbb{B}(\bar{x};\delta).$$

反之, 假设对任意的 $\varepsilon>0$, 存在 $\delta>0$ 使得式 (4.2) 成立. 任取 $\lambda<f(\bar{x})$, 并选取 $\varepsilon>0$ 满足 $\lambda<f(\bar{x})-\varepsilon$, 则对某 $\delta>0$, 有

$$\lambda<f(\bar{x})-\varepsilon<f(x),\quad \forall\, x\in\mathbb{B}(\bar{x};\delta),$$

这就完成了证明. □

命题 4.4 函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是 l.s.c.的当且仅当对任意数 $\lambda\in\mathbb{R}$, 水平集 $\{x\in\mathbb{R}^n\mid f(x)\leqslant\lambda\}$ 是闭的.

证明 假设 f 是 l.s.c.的, 对任意固定的 $\lambda\in\mathbb{R}$, 记

$$\Omega:=\{x\in\mathbb{R}^n\mid f(x)\leqslant\lambda\}.$$

如果 $\bar{x}\notin\Omega$, 有 $f(\bar{x})>\lambda$. 根据定义, 找到 $\delta>0$ 使得

$$f(x)>\lambda,\quad \forall\, x\in\mathbb{B}(\bar{x};\delta),$$

这表明 $\mathbb{B}(\bar{x};\delta)\subset\Omega^c$($\Omega$ 的余集). 因此 Ω^c 是开的, 从而 Ω 是闭的.

为证反向的蕴涵关系成立, 假设对任意的 $\lambda \in \mathbb{R}$, 集合 $\{x \in \mathbb{R}^n \mid f(x) \leqslant \lambda\}$ 是闭的. 固定 $\bar{x} \in \mathbb{R}^n$ 和满足且 $f(\bar{x}) > \lambda$ 的 $\lambda \in \mathbb{R}^n$, 接着设

$$\Theta := \{x \in \mathbb{R}^n \mid f(x) \leqslant \lambda\}.$$

由于 Θ 是闭的且 $\bar{x} \notin \Theta$, 存在 $\delta > 0$ 使得 $\mathbb{B}(\bar{x};\delta) \subset \Theta^c$. 这蕴涵着

$$f(x) > \lambda, \quad \forall x \in \mathbb{B}(\bar{x};\delta),$$

因此完成了命题的证明. □

命题 4.5 函数 $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ 是 l.s.c. 的当且仅当它的上图是闭的.

证明 为证必要性部分, 取定 $(\bar{x}, \lambda) \notin \operatorname{epi} f$, 即 $\lambda < f(\bar{x})$. 对任意的 $\varepsilon > 0$ 满足 $f(\bar{x}) > \lambda + \varepsilon > \lambda$ 的 $\varepsilon > 0$, 存在 $\delta > 0$ 使得 $f(x) > \lambda + \varepsilon$ 对任意的 $x \in \mathbb{B}(\bar{x};\delta)$ 成立. 因此

$$\mathbb{B}(\bar{x};\delta) \times (\lambda - \varepsilon, \lambda + \varepsilon) \subset (\operatorname{epi} f)^c,$$

故集合 $\operatorname{epi} f$ 是闭的.

充分性, 假设 $\operatorname{epi} f$ 是闭的. 利用命题 4.4, 只需证明对任意的 $\lambda \in \mathbb{R}$, 集合 $\Omega := \{x \in \mathbb{R}^n \mid f(x) \leqslant \lambda\}$ 是闭的. 为此, 任取收敛于 $\bar{x}$ 的序列 $\{x_k\} \subset \Omega$. 由于 $f(x_k) \leqslant \lambda$, 对任意的 k 有 $(x_k, \lambda) \in \operatorname{epi} f$. 由 $(x_k, \lambda) \to (\bar{x}, \lambda)$ 得 $(\bar{x}, \lambda) \in \operatorname{epi} f$, 从而 $f(\bar{x}) \leqslant \lambda$. 因此 $\bar{x} \in \Omega$, 这就证明了集合 Ω 的闭性. □

命题 4.6 函数 $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ 在 $\bar{x}$ 是 l.s.c. 的当且仅当对任何序列 $x_k \to \bar{x}$, 有

$$\liminf_{k\to\infty} f(x_k) \geqslant f(\bar{x}).$$

证明 如果 f 在 $\bar{x}$ 处是 l.s.c.的, 取序列 $x_k \to \bar{x}$. 对任意的 $\lambda < f(\bar{x})$, 存在 $\delta > 0$ 使得式 (4.1) 成立. 于是 $x_k \in \mathbb{B}(\bar{x};\delta)$, 从而对大的 k 有 $\lambda < f(x_k)$. 于是 $\lambda \leqslant \liminf\limits_{k\to\infty} f(x_k)$, 因此根据 $\lambda < f(\bar{x})$ 的任意选择性得 $f(\bar{x}) \leqslant \liminf\limits_{k\to\infty} f(x_k)$.

为证反向的结论成立, 假设相反 f 在 $\bar{x}$ 处不是 l.s.c. 的, 则存在 $\lambda < f(\bar{x})$ 使得对任意的 $\delta > 0$, 有 $x_\delta \in \mathbb{B}(\bar{x};\delta)$ 且 $\lambda \geqslant f(x_\delta)$. 把它应用到 $\delta_k := \dfrac{1}{k}$, 得收敛于 $\bar{x}$ 的序列 $\{x_k\} \subset \mathbb{R}^n$ 且 $\lambda \geqslant f(x_k)$. 因此

$$f(\bar{x}) > \lambda \geqslant \liminf_{k\to\infty} f(x_k),$$

这是一个矛盾, 从而完成了证明. □

现在已经准备好得到关于极小值函数 (3.9) 的下半连续性的重要结果, 它在下面是有用的.

定理 4.7 在定理 3.27 的条件下且 $0 \in F$, 则极小时间函数 $\mathcal{T}_\Omega^F$ 是 l.s.c. 的.

证明 取 $\bar{x} \in \mathbb{R}^n$, 并任意取定当 $k \to \infty$ 时收敛于 $\bar{x}$ 的序列 $\{x_k\}$. 根据命题 4.6, 我们只需验证

$$\liminf_{k\to\infty} \mathcal{T}_\Omega^F(x_k) \geqslant \mathcal{T}_\Omega^F(\bar{x}).$$

如果 $\liminf_{k\to\infty} \mathcal{T}_\Omega^F(x_k) = \infty$, 上面的不等式显然成立; 因此余下需考虑 $\liminf_{k\to\infty} \mathcal{T}_\Omega^F(x_k) = \gamma \in [0, \infty)$ 的情形. 由于只需证明 $\gamma \geqslant \mathcal{T}_\Omega^F(x_{\bar{k}})$, 不失一般性, 假设 $\mathcal{T}_\Omega^F(x_k) \to \gamma$. 根据极小时间函数的定义, 找到序列 $\{t_k\} \subset [0, \infty)$ 使得

$$\mathcal{T}_\Omega^F(x_k) \leqslant t_k \leqslant \mathcal{T}_\Omega^F(x_k) + \frac{1}{k}, \quad (x_k + t_k F) \cap \Omega \neq \varnothing, \quad k \in \mathbb{N}.$$

对任意的 $k \in \mathbb{N}$, 取定 $f_k \in F$, $\omega_k \in \Omega$ 满足 $x_k + t_k f_k = \omega_k$, 并不失一般性, 假设 $\omega_k \to \omega \in \Omega$. 考虑两种情形: $\gamma > 0$ 和 $\gamma = 0$. 如果 $\gamma > 0$, 则

$$f_k \to \frac{\omega - \bar{x}}{\gamma} \in F, \quad k \to \infty,$$

这蕴涵着 $(\bar{x} + \gamma F) \cap \Omega \neq \varnothing$, 因此 $\gamma \geqslant \mathcal{T}_\Omega^F(\bar{x})$. 考虑另一种情形 $\gamma = 0$. 此时序列 $\{t_k\}$ 收敛于 0, 因此 $t_k f_k \to \omega - \bar{x}$. 由命题 3.24 得 $\omega - \bar{x} \in F_\infty$, 从而 $\bar{x} \in \Omega - F_\infty$. 利用定理 3.27 知 $\mathcal{T}_\Omega^F(\bar{x}) = 0$, 而且 $\gamma \geqslant \mathcal{T}_\Omega^F(\bar{x})$. 这就在定理 3.27 的条件下证明了 $\mathcal{T}_\Omega^F$ 的下半连续性. □

下面两个命题表明在函数重要运算下下半连续性得以保持的两种情形.

命题 4.8 设 $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}, i = 1, \cdots, m$, 在某点 $\bar{x}$ 处是 l.s.c. 的, 则和 $\sum_{i=1}^m f_i$ 在该点也是 l.s.c. 的.

证明 只需证明当 $m = 2$ 时结论成立. 取 $x_k \to \bar{x}$, 得

$$\liminf_{k\to\infty} f_i(x_k) \geqslant f(\bar{x}), \quad i = 1, 2,$$

即蕴涵着

$$\liminf_{k\to\infty} [f_1(x_k) + f_2(x_k)] \geqslant \liminf_{k\to\infty} f_1(x_k) + \liminf_{k\to\infty} f_2(x_k) \geqslant f_1(\bar{x}) + f_2(\bar{x}),$$

因此根据命题 4.6 可知 $f_1 + f_2$ 在 $\bar{x}$ 处是 l.s.c. 的. □

命题 4.9 设 I 是任意非空指标集, 设 $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ 是 l.s.c. 的, $\forall i \in I$, 则上确界函数 $f(x) := \sup_{i \in I} f_i(x)$ 也是 l.s.c..

证明 对任意的数 $\lambda \in \mathbb{R}$, 由定义直接得

$$\{x \in \mathbb{R}^n \mid f(x) \leqslant \lambda\} = \left\{x \in \mathbb{R}^n \mid \sup_{i \in I} f_i(x) \leqslant \lambda\right\} = \bigcap_{i \in I} \{x \in \mathbb{R}^n \mid f_i(x) \leqslant \lambda\}.$$

因此 f 的任何水平集作为闭集的交是闭的. 根据命题 4.4 保证了 f 的下半连续性.

□

下面的结果给出定理 1.18 中的经典 Weierstrass 存在性定理的相互关联的两个单边的版本. 关键的差别在于先假设下半连续性而不是连续性, 但只保证最小值的存在性, 而不是像定理 1.18 那样也保证最大值的存在性.

定理 4.10 设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是 l.s.c. 函数, 则下列结论成立:

(i) 函数 f 在任何与其有效域相交的非空紧子集 $\Omega\subset\mathbb{R}^n$ 上达到它的绝对极小值.

(ii) 假设 $\inf\{f(x)\mid x\in\mathbb{R}^n\}<\infty$, 假设存在一实数 $\lambda>\inf\{f(x)\mid x\in\mathbb{R}^n\}$ 使得 f 的水平集 $\{x\in\mathbb{R}^n\mid f(x)<\lambda\}$ 是有界的, 则 f 在 $\mathbb{R}^n$ 上的某点 $\bar{x}\in\operatorname{dom}f$ 达到它的绝对极小值.

证明 为证 (i), 先证 f 在 Ω 上是下有界的. 假设, 则对任意的 $k\in\mathbb{N}$, 存在 $x_k\in\Omega$ 使得 $f(x_k)\leqslant -k$. 由于 Ω 是紧的, 不失一般性, 得 $x_k\to\bar{x}\in\Omega$. 由下半连续得

$$f(\bar{x})\leqslant\liminf_{k\to\infty}f(x_k)=-\infty,$$

这是一个矛盾, 从而证明了 f 在 Ω 上是下有界的. 现证断言 (i) 成立, 注意到 $\gamma:=\inf_{x\in\Omega}f(x)<\infty$. 取 $\{x_k\}\subset\Omega$ 使得 $f(x_k)\to\gamma$. 根据 Ω 的紧性, 假设 $x_k\to\bar{x}\in\Omega(k\to\infty)$, 则

$$f(\bar{x})\leqslant\liminf_{k\to\infty}f(x_k)=\gamma,$$

这表明 $f(\bar{x})=\gamma$, 因此 $\bar{x}$ 为 f 在 Ω 上的绝对极小值.

余下证明断言 (ii) 成立. 不难证明 f 是下有界的, 因此 $\gamma:=\inf\limits_{x\in\mathbb{R}^n}f(x)$ 是一实数. 设 $\{x_k\}\subset\mathbb{R}^n$ 满足 $f(x_k)\to\gamma(k\to\infty)$, 并设 $k_0\in\mathbb{N}$ 满足对任意的 $k\geqslant k_0, f(x_k)<\lambda$. 由于水平集 $\{x\in\mathbb{R}^n\mid f(x)<\lambda\}$ 是有界的, 选取收敛于某 $\bar{x}$ 的 $\{x_k\}$ 的子序列 (不再重新标记), 则

$$f(\bar{x})\leqslant\liminf_{k\to\infty}f(x_k)\leqslant\gamma,$$

这表明 $\bar{x}\in\operatorname{dom}f$ 是 f 在 $\mathbb{R}^n$ 上的一个极小值点. □

与本节上面所给的所有结果不同, 下面的定理要求凸性. 它提供了定理 4.10(ii) 中的水平集有界性的有效刻画. 值得注意的是, 下面定理的条件 (iv) 称为强制性.

定理 4.11 设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 是凸的, 设 $\inf\{f(x)\mid x\in\mathbb{R}^n\}<\infty$, 则下面的断言是等价的:

(i) 存在一个实数 $\beta>\inf\{f(x)\mid x\in\mathbb{R}^n\}$ 使得水平集 $\mathcal{L}_\beta:=\{x\in\mathbb{R}^n\mid f(x)<\beta\}$ 是有界的.

(ii) f 的所有水平集 $\{x\in\mathbb{R}^n\mid f(x)<\lambda\}$ 都是有界的.

(iii) $\lim_{\|x\|\to\infty} f(x) = \infty$.

(iv) $\liminf_{\|x\|\to\infty} \dfrac{f(x)}{\|x\|} > 0$.

证明 令 $\gamma := \inf\{f(x) \mid x \in \mathbb{R}^n\}$, 它是一个实数或是 $-\infty$. 先证 (i)⇒(ii). 假设相反, 则存在 $\lambda \in \mathbb{R}$ 使得水平集 $\mathcal{L}_\lambda$ 不是有界的, 则存在序列 $\{x_k\} \subset \mathcal{L}_\lambda$ 使得 $\|x_k\| \to \infty$. 取定 $\alpha \in (\gamma, \beta), \bar{x} \in \mathcal{L}_\alpha$, 定义凸组合

$$y_k := \bar{x} + \frac{1}{\sqrt{\|x_k\|}}(x_k - \bar{x}) = \frac{1}{\sqrt{\|x_k\|}}x_k + \left(1 - \frac{1}{\sqrt{\|x_k\|}}\right)\bar{x}.$$

根据 f 的凸性, 有关系

$$\begin{aligned} f(y_k) &\leqslant \frac{1}{\sqrt{\|x_k\|}} f(x_k) + \left(1 - \frac{1}{\sqrt{\|x_k\|}}\right) f(\bar{x}) \\ &< \frac{\lambda}{\sqrt{\|x_k\|}} + \left(1 - \frac{1}{\sqrt{\|x_k\|}}\right)\alpha \to \alpha < \beta, \end{aligned}$$

因此对充分大的 k, 有 $y_k \in \mathcal{L}_\beta$, 这就确保 $\{y_k\}$ 的有界性. 由三角不等式得

$$\|y_k\| \geqslant \left\|\frac{x_k}{\sqrt{\|x_k\|}}\right\| - \left(1 - \frac{1}{\sqrt{\|x_k\|}}\right)\|\bar{x}\| \geqslant \sqrt{\|x_k\|} - \left(1 - \frac{1}{\sqrt{\|x_k\|}}\right)\|\bar{x}\| \to \infty,$$

这是一个矛盾. 因此 (ii) 成立.

下证 (ii)⇒(iii). 事实上, (iii) 不成立意味着存在一个常数 $M > 0$, 存在一个序列 $\{x_k\}$ 使得 $\|x_k\| \to \infty\,(k \to \infty)$ 且 $f(x_k) < M, \forall k$. 但是这显然与水平集 $\{x \in \mathbb{R}^n \mid f(x) < M\}$ 的有界性矛盾.

最后证明 (iii)⇒(i), 把 (iii)⇔(iv) 的证明留作练习. 假设 (iii) 成立, 对 $M > \gamma$, 找到 α 使得 $f(x) \geqslant M, \forall \|x\| > \alpha$. 于是,

$$\mathcal{L}_m := \{x \in \mathbb{R}^n \mid f(x) < M\} \subset \mathbb{B}(0;\alpha),$$

这表明这个集合是有界的, 因此证明了 (i) 成立. □

4.2 最优性条件

本节首先得到下面集合约束凸最优化问题:

$$\min\ f(x), \quad \text{s.t.}\, x \in \Omega, \tag{4.3}$$

最优解的必要和充分条件, 这里 Ω 是非空凸集, $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ 是凸函数.

定义 4.12 称 $\bar{x} \in \Omega$ 是式 (4.3) 的一个局部最优解, 如果 $f(\bar{x}) < \infty$, 并且存在 $\gamma > 0$ 使得

$$f(x) \geqslant f(\bar{x}), \quad \forall\, x \in \mathbb{B}(\bar{x};\gamma) \cap \Omega. \tag{4.4}$$

如果 $f(x) \geqslant f(\bar{x}), \forall\, x \in \Omega$, 则称 $\bar{x} \in \Omega$ 是式 (4.3) 的一个全局最优解.

特别地，下面的结论表明对凸最优问题, 局部最优解与全局最优解相同.

命题 4.13 下面的断言在式 (4.3) 的凸的情形下是等价的:

(i) $\bar{x}$ 是约束问题 (4.3) 的一个局部最优解.

(ii) $\bar{x}$ 是无约束问题

$$\min g(x) := f(x) + \delta(x;\Omega), \quad x \in \mathbb{R}^n \tag{4.5}$$

的一个局部最优解.

(iii) $\bar{x}$ 是无约束问题 (4.5) 的一个全局最优解.

(iv) $\bar{x}$ 是约束问题 (4.3) 的一个全局最优解.

证明 为证 (i) $\Rightarrow$(ii), 选取 $\gamma > 0$ 使得式 (4.4) 成立. 由于 $g(x) = f(x), \forall\, x \in \Omega \cap \mathbb{B}(\bar{x};\gamma)$, 又由于 $g(x) = \infty$, 若 $x \in \mathbb{B}(\bar{x};\gamma)$ 且 $x \notin \Omega$, 所以

$$g(x) \geqslant g(\bar{x}), \quad \forall\, x \in \mathbb{B}(\bar{x};\gamma),$$

这意味着 (ii) 成立. 蕴涵关系 (ii)$\Rightarrow$(iii) 是命题 2.33 的一个直接推论. (iii)$\Rightarrow$(iv) 的证明类似于 (i)$\Rightarrow$(ii) 的证明, 而 (iv)$\Rightarrow$(i) 是显然的. □

下面对凸最优化问题使用术语“最优解”, 而不再提是全局的或局部的. 值得注意的是, 关联于约束问题 (4.3) 的无约束问题 (4.5) 的目标函数 g 当 $\Omega \neq \mathbb{R}^n$ 时总是增广实值的. 这反映出对违反约束的“无限惩罚”.

下面的定理给出具有关于一般 (凸) 费用函数的约束问题 (4.3) 的最优解的基本必要和充分最优性条件.

定理 4.14 考虑规范条件

$$\partial^{\infty} f(\bar{x}) \cap [-N(\bar{x};\Omega)] = \{0\} \tag{4.6}$$

下的凸问题 (4.3), 其中 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 满足对某点 $\bar{x} \in \Omega, f(\bar{x}) < \infty$. 如果 f 在 $\bar{x}$ 处是连续的, 此规范条件自动成立, 则下面的结论是等价的:

(i) $\bar{x}$ 是问题 (4.3) 的最优解.

(ii) $0 \in \partial f(\bar{x}) + N(\bar{x};\Omega)$, 即 $\partial f(\bar{x}) \cap [-N(\bar{x};\Omega)] \neq \varnothing$.

(iii) $f'(\bar{x};d) \geqslant 0, \forall\, d \in T(\bar{x};\Omega)$.

证明 首先注意到, 如果 f 在 $\bar{x}$ 处是连续的, 则 $\partial^{\infty} f(\bar{x}) = \{0\}$, 从而式 (4.6) 成立. 这由定理 2.29 可得. 现为证 (i)$\Rightarrow$(ii), 根据命题 4.13 应用问题 (4.3) 的无约

束形式 (4.5). 于是命题 2.35 给出了最优解刻画 $0 \in \partial g(\bar{x})$. 在规范条件 (4.6) 下应用定理 2.44 中的次微分和法则即得 (ii) 中的最优性条件.

为证 (ii)⇒(iii), 利用切锤–法锤对偶性 (3.4) 有等价关系:

$$d \in T(\bar{x};\Omega) \Leftrightarrow [\langle v,d\rangle \leqslant 0, \forall v \in N(\bar{x};\Omega)].$$

取定元素 $w \in \partial f(\bar{x})$ 使得 $-w \in N(\bar{x};\Omega)$, 则 $\langle w,d\rangle \geqslant 0, \forall d \in T(\bar{x};\Omega)$, 根据定理 2.84 得 $f'(\bar{x};d) \geqslant \langle w,,d\rangle \geqslant 0$, 由此得 (iii) 成立.

余下证明蕴涵关系 (iii)⇒(i) 成立, 由引理 2.83 得

$$f(\bar{x}+d) - f(\bar{x}) \geqslant f'(\bar{x};d) \geqslant 0, \quad \forall d \in T(\bar{x};\Omega).$$

根据切锥的定义, 有 $T(\bar{x};\Omega) = \mathrm{cl}[\mathrm{cone}(\Omega-\bar{x})]$, 于是当 $w \in \Omega$ 时 $d := w-\bar{x} \in T(\bar{x};\Omega)$. 因此

$$f(w) - f(\bar{x}) = f(w-\bar{x}+\bar{x}) - f(\bar{x}) = f(\bar{x}+d) - f(\bar{x}) \geqslant 0,$$

这就证明了 (i) 成立, 从而完成了定理的证明. □

推论 4.15 假设在定理 4.14 的框架下费用函数 $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ 在 $\bar{x} \in \mathrm{dom}\, f$ 处是可微的, 则下面的结论是等价的:

(i) $\bar{x}$ 是问题 (4.3) 的最优解.

(ii) $-\nabla f(\bar{x}) \in N(\bar{x};\Omega)$, 当 $\bar{x} \in \mathrm{int}\,\Omega$ 时简化为 $\nabla f(\bar{x}) = 0$.

证明 由于凸函数 f 在 $\bar{x}$ 处的可微性蕴涵着它的连续性, 故根据定理 2.29 蕴涵着在该点附近的 Lipschitz 连续性. 由于根据定理 3.3 此时 $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$, 故由定理 4.14 可直接推得该推论成立. □

接下来考虑下面含由有限凸函数给出的函数不等式约束以及如式 (4.3) 中集合/几何约束的凸最优化问题. 简单起见, 假设 f 在 $\mathbb{R}^n$ 上也是有限的, 问题 (P) 是

$$\min f(x), \quad \text{s.t.} \quad g_i(x) \leqslant 0, \quad i = 1,\cdots,m, \ x \in \Omega.$$

定义 (P) 的只涉及费用和函数约束的 Lagrange 函数为

$$L(x,\lambda) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x),$$

其中 $\lambda_i \in \mathbb{R}$, 而活跃指标集由 $I(x) := \{i = \{1,\cdots,m\} \mid g_i(x) = 0\}$ 给出.

定理 4.16 设 $\bar{x}$ 是问题 (P) 的最优解, 则存在不同时为零的乘子 $\lambda_0 \geqslant 0$ 和 $\lambda = (\lambda_1,\cdots,\lambda_m) \in \mathbb{R}^m$, 使得 $\lambda_i \geqslant 0, i = 0,\cdots,m$, 有

$$0 \in \lambda_0 \partial f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \partial g_i(\bar{x}) + N(\bar{x};\Omega),$$

且 $\lambda_i g_i(\bar{x}) = 0, i = 1, \cdots, m$.

证明 考虑有限凸函数

$$\varphi(x) := \max\{f(x) - f(\bar{x}), g_i(x) \mid i = 1, \cdots, m\},$$

则 $\bar{x}$ 是下面无函数约束问题

$$\min \varphi(x), \quad \text{s.t.}\, x \in \Omega$$

的解. 由于 φ 是有限的, 因此是局部 Lipschitz 连续的, 由定理 4.14 得 $0 \in \partial\varphi(\bar{x}) + N(\bar{x}; \Omega)$. 由命题 2.54 的次微分最大值法则得

$$0 \in \text{co}\left[\partial f(\bar{x}) \cup \left[\bigcup_{i \in I(\bar{x})} \partial g_i(\bar{x})\right]\right] + N(\bar{x}; \Omega).$$

因此 $\lambda_0 \geqslant 0, \lambda_i \geqslant 0, i \in I(\bar{x})$ 使得 $\lambda_0 + \sum\limits_{i \in I(\bar{x})} \lambda_i = 1$ 及

$$0 \in \lambda_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \lambda_i \partial g_i(\bar{x}) + N(\bar{x}; \Omega). \tag{4.7}$$

令 $\lambda_i := 0,\ i \notin I(\bar{x})$, 得 $\lambda_i g_i(\bar{x}) = 0,\ i = 1, \cdots, m$ 且 $(\lambda_0, \lambda) \neq 0$. □

许多应用包括数值算法来解问题 (P) 需要有效的约束规范条件保证在定理 4.16 中 $\lambda_0 \neq 0$. 下面的约束规范在凸最优化中是经典的.

定义 4.17 在问题 (P) 中称 Slater 约束规范 (CQ) 成立, 如果存在 $u \in \Omega$ 使得 $g_i(u) < 0, i = 1, \cdots, m$.

接下来的定理给出在规范的、正常的，或 Karush-Kuhn-Tucker(KKT) 形式下凸最优性的必要和充分条件.

定理 4.18 假设在问题 (P) 中 Slater CQ 成立, 则 $\bar{x}$ 是问题 (P) 的最优解当且仅当存在非负 Lagrange 乘子 $(\lambda_1, \cdots, \lambda_m) \in \mathbb{R}^m$ 使得

$$0 \in \partial f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \partial g_i(\bar{x}) + N(\bar{x}; \Omega) \tag{4.8}$$

和 $\lambda_i g_i(\bar{x}) = 0, i = 1, \cdots, m$.

证明 为证必要性部分, 只需证明在式 (4.7) 中 $\lambda_0 \neq 0$. 若相反地假设 $\lambda_0 = 0$, 则可找到 $(\lambda_1, \cdots, \lambda_m) \neq 0, v_i \in \partial g_i(\bar{x}), v \in N(\bar{x}; \Omega)$ 满足

$$0 = \sum_{i=1}^{m} \lambda_i v_i + v.$$

这蕴涵着

$$0 \in \sum_{i=1}^{m} \lambda_i \langle v_i, x - \bar{x} \rangle + \langle v, x - \bar{x} \rangle \leqslant \sum_{i=1}^{m} \lambda_i (g_i(x) - g_i(\bar{x})) = \sum_{i=1}^{m} \lambda_i g_i(x), \quad \forall\, x \in \Omega,$$

矛盾, 因为 Slater 条件蕴涵着 $\sum_{i=1}^{m} \lambda_i g_i(u) < 0$.

为验证式 (4.8) 的充分性, 取 $v_0 \in \partial f(x), v_i \in \partial g_i(\bar{x}), v \in N(\bar{x};\Omega)$ 使得 $0 = v_0 + \sum_{i=1}^{m} \lambda_i v_i + v$ 且 $\lambda_i g_i(\bar{x}) = 0$, $\lambda_i \geqslant 0$, $i = 1, \cdots, m$. 于是问题 (P) 中 $\bar{x}$ 的最优性直接由次微分和法锥的定义可得. 事实上, 对任意的 $x \in \Omega$ 且 $g_i(x) \leqslant 0$, $i = 1, \cdots, m$, 有

$$\begin{aligned} 0 = \langle \lambda_i v_i + v, x - \bar{x} \rangle &= \sum_{i=1}^{m} \lambda_i \langle v_i, x - \bar{x} \rangle + \langle v, x - \bar{x} \rangle \\ &\leqslant \sum_{i=1}^{m} \lambda_i [g_i(x) - g_i(\bar{x})] + f(x) - f(\bar{x}) \\ &= \sum_{i=1}^{m} \lambda_i g_i(x) + f(x) - f(\bar{x}) \leqslant f(x) - f(\bar{x}), \end{aligned}$$

这就完成了证明. □

在本节最后, 考虑具有不等式和线性等式约束的凸最优化问题 (Q):

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & g_i(x) \leqslant 0, \quad i = 1, \cdots, m, \quad Ax = b, \end{aligned}$$

其中 A 为 $p \times n$ 矩阵, $b \in \mathbb{R}^p$.

推论 4.19 假设关于问题 (Q) 的定义 4.17 中的 Slater CQ 成立, 其中集合 $\Omega := \{x \in \mathbb{R}^n \mid Ax - b = 0\}$, 则 $\bar{x}$ 是 (Q) 的最优解当且仅当存在非负乘子 $\lambda_1, \cdots, \lambda_m$ 使得

$$0 \in \partial f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \partial g_i(\bar{x}) + \operatorname{im} A^*$$

和 $\lambda_i g_i(\bar{x}) = 0, i = 1, \cdots, m$.

证明 由定理 4.16 和命题 2.12 即得, 其中集合 Ω 的法锥通过伴随算子 A^* 的像来计算. □

本节最后的结果给出具有可微数据凸最优化的 Lagrange 乘子法则的经典版本.

推论 4.20 设 f 和 g_i, $i = 1, \cdots, m$ 在 $\bar{x}$ 处是可微的, 设 $\{\nabla g_i(\bar{x}) \mid i \in I(\bar{x})\}$ 是线性无关的, 则 $\bar{x}$ 是 $\Omega = \mathbb{R}^n$ 时问题 (P) 的最优解当且仅当存在非负乘子 $\lambda_1, \cdots, \lambda_m$ 使得

$$0 \in \nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) = L_x(\bar{x}, \lambda)$$

和 $\lambda_i g_i(\bar{x}) = 0, i = 1, \cdots, m$.

证明　此时 $N(\bar{x}; \Omega) = \{0\}$, 由式 (4.7) 得, $\lambda_0 = 0$ 与梯度向量 $\{\nabla g_i(\bar{x}) \mid i \in I(\bar{x})\}$ 的线性无关性矛盾. □

4.3　凸最优化中的次梯度方法

本节致力于凸最优化的算法方面, 给出经典次梯度方法的两个版本: 一个是对无约束最优化问题, 另一个是对约束最优化问题. 首先, 研究无约束问题:

$$\min f(x), \quad \text{s.t.}\, x \in \mathbb{R}^n, \tag{4.9}$$

这里 $f: \mathbb{R}^n \to \mathbb{R}$ 是凸函数. 设 $\{\alpha_k\}(k \in \mathbb{N})$ 是正数序列. 由 $\{\alpha_k\}$ 生成的次梯度算法定义如下. 给定初始点 $x_1 \in \mathbb{R}^n$, 考虑迭代过程:

$$x_{k+1} := x_k - \alpha_k v_k, \quad \text{其中} \ \ v_k \in \partial f(x_k), \ k \in \mathbb{N}. \tag{4.10}$$

本节假设问题 (4.9) 有最优解, f 在 $\mathbb{R}^n$ 上是凸的, Lipschitz 连续的. 我们的主要目标是找到条件确保算法的收敛性. 下面把主要结果的证明分为几个命题, 它们本身也具有独立的意义. 第一个命题给出了凸 Lipschitz 函数的一个重要次梯度性质.

命题 4.21　设 $\ell \geqslant 0$ 是 f 在 $\mathbb{R}^n$ 上的一个 Lipschitz 常数. 对任意的 $x \in \mathbb{R}^n$, 有次梯度估计

$$\|v\| \leqslant \ell, \quad \forall\, v \in \partial f(x).$$

证明　取定 $x \in \mathbb{R}^n$, 任取次梯度 $v \in \partial f(x)$. 由定义得

$$\langle v, u - x\rangle \leqslant f(u) - f(x), \quad \forall\, u \in \mathbb{R}^n.$$

由上式再结合 f 的 Lipschitz 性质得

$$\langle v, u - x\rangle \leqslant \ell\|u - x\|, \quad u \in \mathbb{R}^n,$$

相应地, 得

$$\langle v, y + x - x\rangle \leqslant \ell\|y + x - x\|, \ \ \text{即} \ \ \langle v, y\rangle \leqslant \ell\|y\|, y \in \mathbb{R}^n.$$

令 $y_i := v$, 得结论 $\|v\| \leqslant \ell$. □

命题 4.22　设 $\ell \geqslant 0$ 是 f 在 $\mathbb{R}^n$ 上的 Lipschitz 函数. 对由算法 (4.10) 生成的序列 $\{x_k\}$, 有

$$\|x_{k+1} - x\|^2 \leqslant \|x_k - x\|^2 - 2\alpha_k(f(x_k) - f(x)) + \alpha_k^2\ell^2, \quad \forall\, x \in \mathbb{R}^n, k \in \mathbb{N}. \tag{4.11}$$

证明 取定 $x \in \mathbb{R}^n, k \in \mathbb{N}$, 由式 (4.10) 得

$$\|x_{k+1} - x\|^2 = \|x_k - \alpha_k v_k - x\|^2 = \|x_k - x - \alpha_k v_k\|^2$$

$$= \|x_k - x\|^2 - 2\alpha_k \langle v_k, x_k - x\rangle + \alpha_k^2 \|v_k\|^2.$$

由于 $v_k \in \partial f(x)_k$ 且根据命题 4.21, 有 $\|v_k\| \leqslant \ell$, 得

$$\langle v_k, x - x_k\rangle \leqslant f(x) - f(x_k).$$

因此式 (4.11) 成立, 完成了命题的证明. □

下面的命题给出了对算法 (4.10) 中的任意序列 $\{\alpha_k\}$, 问题 (4.9) 的最优值的迭代闭性的估计. 记

$$\overline{V} := \min\{f(x) \mid x \in \mathbb{R}^n\}, \quad V_k := \min\{f(x_1), \cdots, f(x_k)\}.$$

命题 4.23 设 $\ell \geqslant 0$ 是 f 在 $\mathbb{R}^n$ 上的 Lipschitz 常数, 设 S 是式 (4.9) 的最优解的集合, 则对任意的 $k \in \mathbb{N}$, 有

$$0 \leqslant V_k - \overline{V} \leqslant \frac{d(x_1; S)^2 + \ell^2 \sum_{i=1}^{k} \alpha_i^2}{2\sum_{i=1}^{k} \alpha_i}. \tag{4.12}$$

证明 把任意的 $\bar{x} \in S$ 代入估计 (4.11) 得

$$\|x_{k+1} - \bar{x}\|^2 \leqslant \|x_k - \bar{x}\|^2 - 2\alpha_k(f(x_k) - f(\bar{x})) + \alpha_k^2 \ell^2.$$

由 $V_k \leqslant f(x_i), i = 1, \cdots, k$, 且 $\overline{V} = f(\bar{x})$, 得

$$\|x_{k+1} - \bar{x}\|^2 \leqslant \|x_1 - \bar{x}\|^2 - 2\sum_{i=1}^{m} \alpha_i(f(x_i) - f(\bar{x})) + \ell^2 \sum_{i=1}^{k} \alpha_i^2$$

$$\leqslant \|x_1 - \bar{x}\|^2 - 2\sum_{i=1}^{k} \alpha_i(V_k - \overline{V}) + \ell^2 \sum_{i=1}^{k} \alpha_i^2.$$

由 $\|x_{k+1} - \bar{x}\|^2 \geqslant 0$ 得

$$2\sum_{i=1}^{k} \alpha_i(V_k - \overline{V}) \leqslant \|x_1 - \bar{x}\|^2 + \ell^2 \sum_{i=1}^{k} \alpha_i^2.$$

由于对任意的 k, $V_k \geqslant \overline{V}$, 可推得估计

$$0 \leqslant V_k - \overline{V} \leqslant \frac{\|x_1 - \bar{x}\|^2 + \ell^2 \sum_{i=1}^{k} \alpha_i^2}{2 \sum_{i=1}^{k} \alpha_i},$$

由于 $\bar{x} \in S$ 是任意选取的, 因此得式 (4.12) 成立. □

下面的推论证明在常数步长情形下, 为得到最优值合适的误差估计所需的步数.

推论 4.24　假设在命题 4.23 的条件下对任意的 $k \in \mathbb{N}$, 有 $\alpha_k = \varepsilon$, 则存在正常数 C 使得

$$0 \leqslant V_k - \overline{V} < \ell^2 \varepsilon, \quad \forall\, k > C/\varepsilon^2.$$

证明　由式 (4.12) 可得

$$0 \leqslant V_k - \overline{V} \leqslant \frac{a + k\ell^2\varepsilon^2}{2k\varepsilon} = \frac{a}{2k\varepsilon} + \frac{\ell^2\varepsilon}{2}, \text{ 其中 } a := d(x_1; S)^2.$$

由于当 $k > a/(\ell^2\varepsilon^2)$ 时 $a/(2k\varepsilon) + \ell^2\varepsilon/2 < \ell^2\varepsilon$, 所以当 $C := a/\ell^2$ 时结论成立. □

作为命题 4.23 的一个直接推论, 下面得到第一个收敛性结果.

命题 4.25　在命题 4.23 的条件下额外假设

$$\sum_{k=1}^{\infty} \alpha_k = \infty \text{ 且 } \lim_{k \to \infty} \alpha_k = 0, \tag{4.13}$$

则有收敛性 $V_k \to \overline{V}\,(k \to \infty)$.

证明　任给 $\varepsilon > 0$, 选取 $K \in \mathbb{N}$ 使得对任意的 $i \geqslant K$, $\alpha_i \ell < \varepsilon$. 对任意的 $k > K$, 利用式 (4.12) 得估计

$$\begin{aligned}
0 \leqslant V_k - \overline{V} &\leqslant \frac{d(x_1; S)^2 + \ell^2 \sum_{i=1}^{K} \alpha_i^2}{2 \sum_{i=1}^{k} \alpha_i} + \frac{\ell^2 \sum_{i=K}^{k} \alpha_i^2}{2 \sum_{i=1}^{k} \alpha_i} \\
&\leqslant \frac{d(x_1; S)^2 + \ell^2 \sum_{i=1}^{K} \alpha_i^2}{2 \sum_{i=1}^{k} \alpha_i} + \frac{\varepsilon \sum_{i=K}^{k} \alpha_i}{2 \sum_{i=1}^{k} \alpha_i} \\
&\leqslant \frac{d(x_1; S)^2 + \ell^2 \sum_{i=1}^{K} \alpha_i^2}{2 \sum_{i=1}^{k} \alpha_i} + \frac{\varepsilon}{2}.
\end{aligned}$$

由假设 $\sum_{k=1}^{\infty}\alpha_k=\infty$ 得 $\lim_{k\to\infty}\sum_{i=1}^{k}\alpha_i=\infty$, 因此

$$0\leqslant\liminf_{k\to\infty}(V_k-\overline{V})\leqslant\limsup_{k\to\infty}(V_k-\overline{V})\leqslant\varepsilon/2.$$

由于 $\varepsilon>0$ 是任意选取的, 所以 $\lim_{k\to\infty}(V_k-\overline{V})=0$ 或 $V_k\to\overline{V}(k\to\infty)$. □

接下来的结果给出关于次梯度算法 (4.10) 中最优值收敛性的重要信息.

命题 4.26 在命题 4.25 的条件下, 有

$$\liminf_{k\to\infty}f(x_k)=\overline{V}.$$

证明 由于 $\overline{V}$ 是 (4.9) 中的最优值, 对任意的 $k\in\mathbb{N}$, 有 $f(x_k)\geqslant\overline{V}$, 所以

$$\liminf_{k\to\infty}f(x_k)\geqslant\overline{V}.$$

为证明相反的不等式成立, 假设相反 $\liminf_{k-\infty}f(x_k)>\overline{V}$, 于是找到一个足够小的数 $\varepsilon>0$ 使得

$$\liminf_{k\to\infty}f(x_k)-2\varepsilon>\overline{V}=\min_{x\in\mathbb{R}^n}f(x)=f(\bar{x}),$$

这里 $\bar{x}$ 是问题 (4.9) 的一个最优解. 因此

$$\liminf_{k\to\infty}f(x_k)-\varepsilon>f(\bar{x})+\varepsilon.$$

当 $\liminf_{k\to\infty}f(x_k)<\infty$ 时, 根据 lim inf 的定义, 存在 $k_0\in\mathbb{N}$ 使得对任意的 $k\geqslant k_0$, 有

$$f(x_k)>\liminf_{k\to\infty}f(x_k)-\varepsilon,$$

因此对任意的 $k\geqslant k_0, f(x_k)-f(\bar{x})>\varepsilon$. 注意到如果 $\liminf_{k\to\infty}f(x_k)=\infty$, 这个结论也成立. 在式 (4.11) 中令 $x=\bar{x}$ 得

$$\|x_{k+1}-\bar{x}\|^2\leqslant\|x_k-\bar{x}\|^2-2\alpha_k(f(x_k)-f(\bar{x}))+\alpha_k^2\ell^2,\quad\forall k\geqslant k_0.\tag{4.14}$$

设 k_1 满足对任意的 $k\geqslant k_1,\alpha_k\ell^2<\varepsilon$, 令 $\bar{k}_1:=\max\{k_1,k_0\}$. 对任意的 $k>\bar{k}_1$, 由估计式 (4.14) 易得

$$\|x_{k+1}-\bar{x}\|^2\leqslant\|x_k-\bar{x}\|^2-\alpha_k\varepsilon\leqslant\cdots\leqslant\|x_{\bar{k}_1}-\bar{x}\|^2-\varepsilon\sum_{j=\bar{k}_1}^{k}\alpha_j,$$

这与式 (4.13) 中的假设 $\sum_{k=1}^{\infty}\alpha_k=\infty$ 矛盾, 从而完成了证明. □

下面将证明次梯度算法 (4.10) 的收敛性成立.

定理 4.27　设式 (4.10) 中的生成序列 $\{\alpha_k\}$ 满足

$$\sum_{k=1}^{\infty} \alpha_k = \infty \text{ 且 } \sum_{k=1}^{\infty} \alpha_k^2 < \infty, \tag{4.15}$$

则序列 $\{V_k\}$ 收敛于最优值 $\overline{V}$, 且式 (4.10) 中的迭代序列 $\{x_k\}$ 收敛于问题 (4.9) 的某最优解 $\bar{x}$.

证明　由于式 (4.15) 蕴涵着式 (4.13) 成立, 根据命题 4.25 序列 $\{V_k\}$ 收敛于 $\overline{V}$. 余下需证 $\{x_k\}$ 收敛于问题 (4.9) 的某最优解 $\bar{x}$. 设 $\ell \geqslant 0$ 是 f 在 $\mathbb{R}^n$ 上的 Lipschitz 常数. 由于问题 (4.9) 的最优解集合 S 假定为非空的, 任取 $\tilde{x} \in S$, 在命题 4.22 的估计 (4.11) 中代入 $x = \tilde{x}$, 得

$$\|x_{k+1} - \tilde{x}\|^2 \leqslant \|x_k - \tilde{x}\|^2 - 2\alpha_k(f(x_k) - \overline{V}) + \alpha_k^2 \ell^2, \quad \forall\, k \in \mathbb{N}. \tag{4.16}$$

由于 $f(x_k) - \overline{V} \geqslant 0$, 由不等式 (4.16) 得

$$\|x_{k+1} - \tilde{x}\|^2 \leqslant \|x_k - \tilde{x}\|^2 + \alpha_k^2 \ell^2,$$

则由数学归纳法推得

$$\|x_{k+1} - \tilde{x}\|^2 \leqslant \|x_k - \tilde{x}\|^2 + \ell^2 \sum_{i=1}^{k} \alpha_i^2. \tag{4.17}$$

根据假设 $\sum_{i=1}^{\infty} \alpha_i^2 < \infty$, 式 (4.17) 的右边是有限的, 因此 $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$, 从而

$$\sup_{k \in \mathbb{N}} \|x_{k+1} - \tilde{x}\| < \infty.$$

这蕴涵着序列 $\{x_k\}$ 是有界的. 而且, 由命题 4.26 得

$$\liminf_{k \to \infty} f(x_k) = \overline{V}.$$

设 $\{x_{k_j}\}$ 是 $\{x_k\}$ 的子序列满足

$$\lim_{j \to \infty} f(x_{k_j}) = \overline{V}, \tag{4.18}$$

则 $\{x_{k_j}\}$ 也是有界的, 且有一个极限点 $\bar{x}$. 不失一般性, 假设当 $j \to \infty$ 时 $x_{k_j} \to \bar{x}$. 由式 (4.18) 和 f 的连续性直接推得 $\bar{x}$ 是问题 (4.9) 的最优解. 任取 $j \in \mathbb{N}$ 和 $k \geqslant k_j$. 于是, 当 $\tilde{x} = \bar{x}$ 时式 (4.16) 可重写为

$$\|x_{k+1} - \bar{x}\|^2 \leqslant \|x_k - \bar{x}\|^2 + \alpha_k^2 \ell^2 \leqslant \cdots \leqslant \|x_{k_j} - \bar{x}\|^2 + \ell^2 \sum_{i=k_j}^{k} \alpha_i^2.$$

先对 $k \to \infty$ 时取极限, 然后在所得不等式中, 取当 $j \to \infty$ 时的极限, 得估计

$$\limsup_{k\to\infty} \|x_{k+1} - \bar{x}\|^2 \leqslant \lim_{j\to\infty} \|x_{k_j} - \bar{x}\|^2 + \ell^2 \lim_{j\to\infty} \sum_{i=k_j}^{\infty} \alpha_i^2.$$

由于 $x_{k_j} \to \bar{x}$ 且 $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$, 这蕴涵着

$$\limsup_{k\to\infty} \|x_{k+1} - \bar{x}\|^2 = 0,$$

因此证明了所断言的收敛性成立, 即当 $k \to \infty$ 时, $x_k \to \bar{x}$. □

本节最后, 导出由

$$\min f(x), \quad \text{s.t.} x \in \Omega \tag{4.19}$$

给出的约束最优化凸问题的次梯度算法版本, 这里 $f : \mathbb{R}^n \to \mathbb{R}$ 是凸函数且在 $\mathbb{R}^n$ 上是 Lipschitz 连续的, 具有某常数 $\ell \geqslant 0$, 然而与问题 (4.9) 不同, 现有关于 x 的由非空闭凸集 $\Omega \subset \mathbb{R}^n$ 给出的约束. 下面关于约束问题的次梯度算法是 (4.10) 的变种, 称为投影次梯度方法.

给定正数序列 $\{\alpha_k\}$, 给定初始点 $x_1 \in \Omega$, 迭代序列 $\{x_k\}$ 由

$$x_{k+1} := \Pi(x_k - \alpha_k v_k; \Omega), \quad \text{其中} \quad v_k \in \partial f(x_k),\ k \in \mathbb{N} \tag{4.20}$$

构造, 这里 $\Pi(x; \Omega)$ 表示 x 到 Ω 上的 (唯一)Euclid 投影.

下面的命题, 是对算法 (4.20) 的命题 4.22 的变种, 在证明投影次梯度方法的收敛性结果中起着重要的作用.

命题 4.28 在算法 (4.20) 的一般情形下, 有估计

$$\|x_{k+1} - x\|^2 \leqslant \|x_k - x\|^2 - 2\alpha_k(f(x_k) - f(x)) + \alpha_k^2 \ell^2, \quad \forall x \in \Omega,\ k \in \mathbb{N}.$$

证明 定义 $z_{k+1} := x_k - \alpha_k v_k$, 任取定 $x \in \Omega$, 由命题 4.22 得

$$\|z_{k+1} - x\|^2 \leqslant \|x_k - x\|^2 - 2\alpha_k(f(x_k) - f(x)) + \alpha_k^2 \ell^2, \quad \forall k \in \mathbb{N}. \tag{4.21}$$

现利用 z_{k+1} 在 Ω 上的投影和命题 1.79, 得

$$\|x_{k+1} - x\| = \|\Pi(z_{k+1}; \Omega) - \Pi(x; \Omega)\| \leqslant \|z_{k+1} - x\|.$$

上式结合估计 (4.21) 有

$$\|x_{k+1} - x\|^2 \leqslant \|x_k - x\|^2 - 2\alpha_k(f(x_k) - f(x)) + \alpha_k^2 \ell^2, \quad \forall k \in \mathbb{N}.$$

从而完成了命题的证明. □

基于这个命题, 利用上面次梯度算法 (4.10) 相同的方法, 在关于生成序列 $\{\alpha_k\}$ 的假设 (4.13) 下可得到投影次梯度算法 (4.20) 类似的收敛性结果.

4.4 Fermat-Torricelli 问题

在 1643 年 Pierre de Fermat 向 Evangelista Torricelli 提出下面的最优化问题.

在平面上给定三个点, 找到另外一点使得它与三个给定点的距离之和是最小的. 这个问题由 Torricelli 解决, 被称为 Fermat-Torricelli 问题. Fermat-Torricelli 问题的更一般版本是寻找一点到 $\mathbb{R}^n$ 中有限多个给定点的距离之和是最小的. 这是选址科学中最重要的问题之一, 已被许多学者研究. 解一般 Fermat-Torricelli 问题的第一个数值算法由 Endre Weiszfeld 于 1937 年引入. 保证 Weiszfeld 算法收敛性的假设由 Harold Kuhn 于 1972 年给出. Kuhn 也举出一个例子, 在该例中 Weiszfeld 算法不收敛. 最近几个新算法的引入改进了 Weiszfeld 算法. 本节的目的是利用凸分析和最优化的技术, 从理论和数值的角度再访 Fermat-Torricelli 问题.

给定点 $a_i \in \mathbb{R}^n$, $i = 1, \cdots, m$, 定义 (非光滑) 凸函数

$$\varphi(x) := \sum_{i=1}^{n} \|x - a_i\|, \tag{4.22}$$

则 Fermat-Torricelli 问题的数学描述为

$$\min \varphi(x), \quad \forall x \in \mathbb{R}^n. \tag{4.23}$$

命题 4.29　Fermat-Torricelli 问题 (4.23) 至少有一个解.

证明　结论由定理 4.10 直接可得. □

命题 4.30　假设 a_i, $i = 1, \cdots, m$ 不共线, 即它们不在同一条直线上, 则式 (4.22) 中的函数 φ 是严格凸的, 从而 Fermat-Torricelli 问题 (4.23) 有唯一解.

证明　由于每个函数 $\varphi_i(x) := \|x - a_i\| (i = 1, \cdots, m)$ 显然是凸的, 它们的和 $\varphi = \sum_{i=1}^{m} \varphi_i$ 也是凸的, 即对任意的 $x, y \in \mathbb{R}^n$, 任意的 $\lambda \in (0, 1)$, 有

$$\varphi(\lambda x + (1 - \lambda)y) \leqslant \lambda\varphi(x) + (1 - \lambda)\varphi(y). \tag{4.24}$$

若设 φ 不是严格凸的, 则可找到 $\bar{x}, \bar{y} \in \mathbb{R}^n$ 且 $\bar{x} \neq \bar{y}, \lambda \in (0, 1)$ 使得式 (4.24) 作为等式成立. 于是

$$\varphi_i(\lambda\bar{x} + (1 - \lambda)\bar{y}) = \lambda\varphi_i(\bar{x}) + (1 - \lambda)\varphi_i(\bar{y}), \quad i = 1, \cdots, m,$$

因此保证

$$\|\lambda(\bar{x} - a_i) + (1 - \lambda)(\bar{y} - a_i)\| = \|\lambda(\bar{x} - a_i)\| + \|(1 - \lambda)(\bar{y} - a_i)\|, \quad i = 1, \cdots, m.$$

如果 $\bar{x} \neq a_i$, 且 $\bar{y} \neq a_i$, 则存在 $t_i > 0$ 使得 $t_i\lambda(\bar{x} - a_i) = (1-\lambda)(\bar{y} - a_i)$, 因此

$$\bar{x} - a_i = \gamma_i(\bar{y} - a_i), \quad \text{其中} \quad \gamma_i := \frac{1-\lambda}{t_i\lambda}.$$

由于 $\bar{x} \neq \bar{y}$, 有 $\gamma_i \neq 1$. 于是

$$a_i = \frac{1}{1-\gamma_i}\bar{x} - \frac{\gamma_i}{1-\gamma_i}\bar{y} \in \mathcal{L}(\bar{x}, \bar{y}),$$

这里 $\mathcal{L}(\bar{x}, \bar{y})$ 表示连接 $\bar{x}$ 和 $\bar{y}$ 的直线. 当 $\bar{x} = a_i$ 和 $\bar{y} = a_i$ 两种情形都有 $a_i \in \mathcal{L}(\bar{x}, \bar{y})$. 因此 $a_i \in \mathcal{L}(\bar{x}, \bar{y}), i = 1, \cdots, m$, 矛盾. □

接下来我们进一步解决关于 $\mathbb{R}^n$ 中三个点的 Fermat-Torricelli 问题. 首先给出两个基本引理.

引理 4.31 设 a_1 和 a_2 是 $\mathbb{R}^n$ 中任意两个单位向量, 则有 $-a_1 - a_2 \in \mathbb{B}$ 当且仅当 $\langle a_1, a_2 \rangle \leqslant -\dfrac{1}{2}$.

证明 如果 $-a_1 - a_2 \in \mathbb{B}$, 则 $a_1 + a_2 \in \mathbb{B}$, 且

$$\|a_1 + a_2\| \leqslant 1.$$

两边平方, 得不等式

$$\|a_1 + a_i\|^2 \leqslant 1,$$

于是得

$$\|a_1\|^2 + 2\langle a_1, a_2 \rangle + \|a_2\|^2 = 2 + 2\langle a_1, a_2 \rangle \leqslant 1.$$

因此, $\langle a_1, a_2 \rangle \leqslant -\dfrac{1}{2}$. 反向蕴涵关系的证明是类似的. □

引理 4.32 设 $a_1, a_2, a_3 \in \mathbb{R}^n$ 且 $\|a_1\| = \|a_2\| = \|a_3\| = 1$, 则 $a_1 + a_2 + a_3 = 0$ 当且仅当有等式

$$\langle a_1, a_2 \rangle = \langle a_2, a_3 \rangle = \langle a_3, a_1 \rangle = -\frac{1}{2}. \tag{4.25}$$

证明 由条件 $a_1 + a_2 + a_3 = 0$ 得

$$\|a_1\|^2 + \langle a_1, a_2 \rangle + \langle a_1, a_3 \rangle = 0, \tag{4.26}$$

$$\langle a_2, a_1 \rangle + \|a_2\|^2 + \langle a_2, a_3 \rangle = 0, \tag{4.27}$$

$$\langle a_3, a_1 \rangle + \langle a_3, a_2 \rangle + \|a_3\|^2 = 0. \tag{4.28}$$

式 (4.28) 减去式 (4.26), 得 $\langle a_1, a_2 \rangle = \langle a_2, a_3 \rangle$, 类似地, $\langle a_2, a_3 \rangle = \langle a_3, a_1 \rangle$. 把这些等式代入式 (4.26)~ 式 (4.28) 则得式 (4.25) 成立.

反之, 假设式 (4.25) 成立, 则有

$$\|a_1+a_2+a_3\|^2=\sum_{i=1}^{3}\|a_i\|^2+\sum_{i,j=1,i\neq j}^{3}\langle a_i,a_j\rangle=0,$$

因此有等式 $a_1+a_2+a_3=0$, 从而完成了证明. □

下面利用凸函数的次微分来解关于 $\mathbb{R}^n$ 中三点的 Fermat-Torricelli 问题. 给定两个非零向量 $u,v\in\mathbb{R}^n$, 定义

$$\cos(u,v):=\frac{\langle u,v\rangle}{\|u\|\cdot\|v\|},$$

并且还考虑, 当 $\bar{x}\neq a_i$ 时的三个向量

$$v_i=\frac{\bar{x}-a_i}{\|\bar{x}-a_i\|},\quad i=1,2,3.$$

几何上, 每个向量 v_i 是单位向量, 方向从顶点 a_i 指向 $\bar{x}$. 注意到经典的 (三点)Fermat-Torricelli 问题总有唯一解, 即使给定点属于同一直线. 事实上, 在这种情形下中间点就是问题的解.

下面的定理充分描述了在两种不同情形下关于三点 Fermat-Torricelli 问题的最优解.

定理 4.33 对关于三点 $a_1,a_2,a_3\in\mathbb{R}^n$ 的 Fermat-Torricelli 问题 (4.23), 下面的论断成立:

(i) 设 $\bar{x}\notin\{a_1,a_2,a_3\}$, 则 $\bar{x}$ 是问题的解当且仅当

$$\cos(v_1,v_2)=\cos(v_2,v_3)=\cos(v_3,v_1)=-1/2.$$

(ii) 设 $\bar{x}\in\{a_1,a_2,a_3\}$, 称 $\bar{x}=a_1$, 则 $\bar{x}$ 是问题的解当且仅当

$$\cos(v_2,v_3)\leqslant-1/2.$$

证明 在情形 (i) 下函数 (4.22) 在 $\bar{x}$ 处是可微的, 则 $\bar{x}$ 是问题 (4.23) 的解当且仅当

$$\nabla\varphi(\bar{x})=v_1+v_2+v_3=0.$$

根据引理 4.32 可知上式成立当且仅当

$$\langle v_i,v_j\rangle=\cos(v_i,v_j)=-1/2,\quad\forall\, i,j\in\{1,2,3\}\ \text{且}\ i\neq j.$$

为证 (ii) 成立, 利用次微分 Fermat 法则 (2.35) 和推论 2.45 中的法则, 有 $\bar{x}=a_1$ 是 Fermat-Torricelli 问题的解当且仅当

$$0\in\partial\varphi(a_1)=v_2+v_3+\mathbb{B}.$$

由于 v_2 和 v_3 是单位向量, 由引理 4.31 得

$$\langle v_2, v_3 \rangle = \cos(v_2, v_3) \leqslant -1/2,$$

这就完成了定理的证明. □

例 4.34 从几何上说明平面上 Fermat-Torricelli 问题解的构造. 考虑问题 (4.23), 其中给定的 $\mathbb{R}^2$ 上的三个点 A, B 和 C, 如图 4.2 所示. 如果三角形 ABC 的一个角大于或等于 120°, 那么根据定理 4.33(ii) 相应的顶点是问题 (4.23) 的解. 考虑三角形三个角都小于 120° 的情形. 如图 4.2 所示, 构造两个等边三角形 ABD 和 ACE, 设 S 是 DC 与 BE 的交点. 两个四边形 $ADBC$ 和 $ABCE$ 是凸的, 因此 S 位于三角形 ABC 的内部. 显然两个三角形 DAC 和 BAE 是全等的. 绕 A 旋转 60° 将三角形 DAC 转到三角形 BAE. 将 CD 转到 BE. 因此 $\angle DSB = 60°$. 设 T 是 S 在这个旋转下的像, 则 T 属于 BE. 因此 $\angle AST = \angle ASE = 60°$. 而且 $\angle DSA = 60°$, 从而 $\angle BSA = 120°$. 现显然 $\angle ASC = 120°$, 而且 $\angle BSC = 120°$. 根据定理 4.33(i), 点 S 是所考虑问题的解.

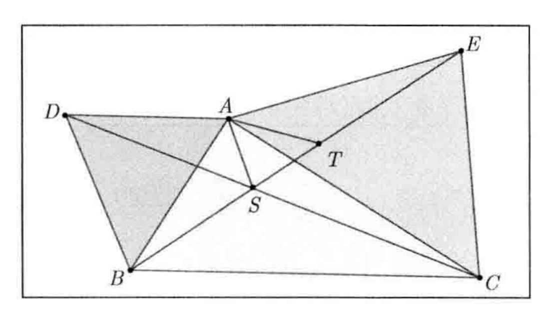


图 4.2 Torricelli 点的构造

下面给出 Weiszfeld 算法[32] 从数值上来解一般情形下关于 $\mathbb{R}^n$ 中的不共线的 m 个点的 Fermat-Torricelli 问题 (4.23). 本节剩余部分总是假设这些点是不共线的. 注意到式 (4.22) 中函数 φ 的梯度可如下计算为

$$\nabla\varphi(x) = \sum_{i=1}^{m} \frac{x - a_i}{\|x - a_i\|}, \quad x \notin \{a_1, \cdots, a_m\}.$$

解梯度方程 $\nabla\varphi(x) = 0$ 得

$$x = \frac{\displaystyle\sum_{i=1}^{m} \frac{a_i}{\|x - a_i\|}}{\displaystyle\sum_{i=1}^{m} \frac{1}{\|x - a_i\|}} =: f(x). \tag{4.29}$$

为在全空间 $\mathbb{R}^n$ 上定义 $f(x)$, 令 $f(x) := x$, 若$x \in \{a_1, \cdots, a_m\}$.

Weiszfeld 算法包括选取任意初始点 $x_1 \in \mathbb{R}^n$, 取 f 如式 (4.29), 循环地定义

$$x_{k+1} := f(x_k), \quad k \in \mathbb{N}, \tag{4.30}$$

其中 $f(x)$ 称为算法映射. 本节余下的部分给出 Weiszfeld 算法收敛性的精确条件及证明, 这是由 Kuhn 在文献 [10] 中提出的, 该文献也纠正了 [32] 中的一些错误. 与文献 [10] 不同的是, 下面将利用凸分析的次微分法则, 这简化了收敛性的证明.

下面的命题可保证每步迭代后 φ 的函数值减少, 即算法的下降性质.

命题 4.35　如果 $f(x) \neq x$, 则 $\varphi(f(x)) < \varphi(x)$.

证明　由假设 $f(x) \neq x$ 知 x 不是 $\{a_1, \cdots, a_m\}$ 中的点. 还注意到点 $f(x)$ 是严格凸函数

$$g(z) := \sum_{i=1}^{m} \frac{\|z - a_i\|^2}{\|x - a_i\|}$$

的唯一最小值点, 而且 $g(f(x)) < g(x) = \varphi(x)$(由于 $f(x) \neq x$). 进一步, 有

$$\begin{aligned} g(f(x)) &= \sum_{i=1}^{m} \frac{\|f(x) - a_i\|^2}{\|x - a_i\|} \\ &= \sum_{i=1}^{m} \frac{(\|x - a_i\| + \|f(x) - a_i\| - \|x - a_i\|)^2}{\|x - a_i\|} \\ &= \varphi(x) + 2(\varphi(f(x)) - \varphi(x)) + \sum_{i=1}^{m} \frac{(\|f(x) - a_i\| - \|x - a_i\|)^2}{\|x - a_i\|}. \end{aligned}$$

这显然蕴涵着不等式

$$2\varphi(f(x)) + \sum_{i=1}^{m} \frac{(\|f(x) - a_i\| - \|x - a_i\|)^2}{\|x - a_i\|} < 2\varphi(x),$$

从而完成了命题的证明. □

下面两个命题表明算法映射 f 在一个顶点附近的性态, 若这个顶点是 Fermat-Torricelli 问题的解. 首先给出一个保证问题 (4.23) 中的顶点的最优性的必要和充分条件. 定义

$$R_j := \sum_{i=1, i \neq j}^{m} \frac{a_i - a_j}{\|a_i - a_j\|}, \quad j = 1, \cdots, m.$$

命题 4.36　顶点 a_j 是问题 (4.23) 的解当且仅当 $\|R_j\| \leqslant 1$.

证明　利用命题 2.35 中的 Fermat 法则和推论 2.45 中的次微分和法则得顶点 a_j 是问题 (4.23) 的解当且仅当

$$0 \in \partial\varphi(a_j) = -R_j + \mathbb{B}(0, 1),$$

显然, 这等价于条件 $\|R_j\| \leqslant 1$. □

来自凸分析的命题 4.36 将本质地简化由 Kuhn 建立的如下证明. 下面使用记号:

$$f^s(x) := f(f^{s-1}(x)), \quad s \in \mathbb{N} \text{ 且 } f^0(x) := x, \quad x \in \mathbb{R}^n.$$

命题 4.37 假设顶点 a_j 不是问题 (4.23) 的解, 则存在 $\delta > 0$ 使得条件 $0 \leqslant \|x - a_j\| \leqslant \delta$ 蕴涵着存在 $s \in \mathbb{N}$ 满足

$$\|f^s(x) - a_j\| > \delta \text{ 且 } \|f^{s-1}(x) - a_j\| \leqslant \delta. \tag{4.31}$$

证明 对任意的 $x \notin \{a_1, \cdots, a_m\}$, 由式 (4.29) 得

$$f(x) - a_j = \frac{\displaystyle\sum_{i=1, i\neq j}^{m} \frac{a_i - a_j}{\|x - a_i\|}}{\displaystyle\sum_{i=1}^{m} \frac{1}{\|x - a_i\|}}.$$

因此

$$\lim_{x \to a_j} \frac{f(x) - a_j}{\|x - a_j\|} = \lim_{x \to a_j} \frac{\displaystyle\sum_{i=1, i\neq j}^{m} \frac{a_i - a_j}{|x - a_i\|}}{1 + \displaystyle\sum_{i=1, i\neq j}^{m} \frac{\|x - a_j\|}{\|x - a_i\|}} = R_j, \quad i = 1, \cdots, m.$$

由命题 4.36 得

$$\lim_{x \to a_j} \frac{\|f(x) - a_j\|}{\|x - a_j\|} = \|R_j\| > 1, \quad j = 1, \cdots, m, \tag{4.32}$$

从而可选取 $\varepsilon > 0, \delta > 0$ 使得

$$\frac{\|f(x) - a_j\|}{\|x - a_j\|} > 1 + \varepsilon, \quad \text{对任意的 } 0 < \|x - a_j\| \leqslant \delta \text{成立}.$$

因此对任意的 $s \in \mathbb{N}$, 对任意的 $q = 1, \cdots, s$, $0 < \|f^{q-1}(x) - a_j\| \leqslant \delta$ 成立使得

$$\|f^s(x) - a_j\| \geqslant (1 + \varepsilon)\|f^{s-1}(x) - a_j\| \geqslant \cdots \geqslant (1 + \varepsilon)^s \|x - a_j\|.$$

由于当 $s \to \infty$ 时 $(1 + \varepsilon)^s \|x - a_j\| \to \infty$, 因此式 (4.31) 成立, 从而完成了证明. □

最后给出 Weiszfeld 算法收敛性的主要结果.

定理 4.38 设 $\{x_k\}$ 是由 Weiszfeld 算法 (4.30) 生成的序列, 设 $x_k \notin \{a_1, \cdots, a_m\}, k \in \mathbb{N}$, 则 $\{x_k\}$ 收敛于问题 (4.32) 的唯一解 $\bar{x}$.

证明　如果对某 $k=k_0, x_k=x_{k+1}$, 则当 $k \geqslant k_0$ 时 $\{x_k\}$ 是一常序列, 因此它收敛于 x_{k_0}. 由于 $f(x_{k_0})=x_{k_0}$ 且 x_{k_0} 不是顶点, 所以 x_{k_0} 是问题的解. 于是可假设对任意的 $k, x_{k+1} \neq x_k$. 由命题 4.35 得非负数列 $\{\varphi(x_k)\}$ 是递减的, 因此它收敛. 从而

$$\lim_{k\to\infty}(\varphi(x_k)-\varphi(x_{k+1}))=0. \tag{4.33}$$

根据 $\{x_k\}$ 的构造知每个 x_k 属于紧集 $\operatorname{co}\{a_1,\cdots,a_m\}$. 因此可从 $\{x_k\}$ 中选取当 $\nu\to\infty$ 时收敛于某点 $\bar{z}$ 的子序列 $\{x_{k_\nu}\}$. 下面将证明 $\bar{z}=\bar{x}$, 即它是问题 (4.23) 的唯一解. 事实上, 由式 (4.33) 得

$$\lim_{\nu\to\infty}(\varphi(x_{k_\nu})-\varphi(f(x_{k_\nu})))=0,$$

由函数 φ 和 f 的连续性知上式蕴涵着 $\varphi(\bar{z})=\varphi(f(\bar{z}))$, 即 $f(\bar{z})=\bar{z}$, 从而如果 $\bar{z}$ 不是顶点, 那么它是问题 (4.23) 的解.

余下考虑 $\bar{z}$ 是顶点的情形, 比如假设 $\bar{z}=a_1$. 相反地, 若 $\bar{z}=a_1\neq\bar{x}$, 则选取充分小的 $\delta>0$ 使得对某 $s\in\mathbb{N}$, 式 (4.31) 成立并且 $\mathbb{B}(a_1;\delta)$ 不包含 $\bar{x}$ 和 $a_i, i=2,\cdots,m$. 由于 $x_{k_\nu}\to\bar{z}=a_1$, 不失一般性, 可假设这个序列全包含在 $\mathbb{B}(a_1;\delta)$ 中. 对 $x=x_{k_1}$, 选取 l_1 使得 $x_{l_1}\in\mathbb{B}(a_1;\delta)$ 且 $f(x_{l_1})\notin\mathbb{B}(a_1;\delta)$. 现选取 $k_\nu>\ell_1$ 并且利用命题 4.37 得 $l_2>l_1$, 且 $x_{l_2}\in\mathbb{B}(a_1;\delta), f(x_{l_2})\notin\mathbb{B}(a_1;\delta)$. 重复这个过程, 找到子序列 $\{x_{l_\nu}\}$ 使得 $x_{l_\nu}\in\mathbb{B}(a_1;\delta)$, 而 $f(x_{l_\nu})$ 不在这个球里. 如果需要抽取另外一个子序列, 可假设 $\{x_{\ell_\nu}\}$ 收敛于某点 $\bar{q}$ 满足 $f(\bar{q})=\bar{q}$. 如果 $\bar{q}$ 不是顶点, 则由上述知它是问题 (4.23) 的解, 因为解 $\bar{x}$ 不在 $\mathbb{B}(a_1;\delta)$ 内, 矛盾. 因此 $\bar{q}$ 是顶点, 且由于其他顶点也不在这个球里它必为 a_1. 于是有

$$\lim_{\nu\to\infty}\frac{\|f(x_{l_\nu})-a_1\|}{\|x_{l_\nu}-a_1\|}=\infty,$$

与命题 4.73 中的式 (4.32) 矛盾, 从而完成了定理的证明. □

4.5　一个广义的 Fermat-Torricelli 问题

在文献中以不同的名称有几个 Fermat-Torricelli 问题的广义版本. 例如, 文献 [15]~[17] 及其里面的参考文献. 值得一提的是下面问题的约束版本也称为“广义 Heron 问题”, 该名称源于古代 Alexandria 城的数学家 Heron, 他在公元 1 世纪时陈述了该问题的一个简化版本. 根据文献 [15], [16], 这里研究一个广义的 Fermat-Torricelli 问题, 它由一个具有 $\mathbb{R}^n$ 中有限个闭凸目标集的极小时间函数 (3.9) 定义, 并具有相同结构的约束集. 在文献 [6], [11], [29] 及其文献中, 读者可找到该问题的

一些特殊版本及选址问题的相关材料. 我们也推荐读者参阅文献 [15], [17], 其中涉及了这类问题的非凸情形.

设目标集合 $\Omega_i, i=1,\cdots,m$ 和约束集合 Ω_0 是 $\mathbb{R}^n$ 中的非空闭凸子集. 对 $\Omega_i, i=1,\cdots,m$, 考虑式 (3.9) 中的极小时间函数 $\mathcal{T}_F(x;\Omega_i)=\mathcal{T}_{\Omega_i}^F(x)$, 其中常数动态学由包含原点作为内点的闭有界凸集 $F\subset\mathbb{R}^n$ 定义, 则广义 Fermat-Torricelli 问题可描述如下

$$\min \mathcal{H}(x) := \sum_{i=1}^{m} \mathcal{T}_F(x;\Omega_i), \quad \text{s.t.}\ x\in\Omega_0. \tag{4.34}$$

本节首要目标是建立最优化问题 (4.34) 解的存在性和唯一性. 下面把证明分为几个本身具有独立意义的命题. 下面从式 (3.11) 中 Minkowski 度规的如下性质开始.

命题 4.39 假设 F 是 $\mathbb{R}^n$ 的闭有界凸子集, 满足 $0\in \operatorname{int}F$, 则 $\rho_F(x)=1$ 当且仅当 $x\in \operatorname{bd}F$.

证明 假设 $\rho_F(x)=1$, 则存在实数序列 $\{t_k\}$ 收敛于 $t=1$, 且对任意的 k 满足 $x\in t_kF$. 由于 F 是闭的, 有 $x\in F$. 为完成充分性的证明, 下证 $x\notin \operatorname{int}F$. 反证法, 假设 $x\in\operatorname{int}F$, 并且选取足够小的 $t>0$ 使得 $x+tx\in F$, 则 $x\in(1+t)^{-1}F$. 从而 $\rho_F(x)\leqslant(1+t)^{-1}<1$, 矛盾.

现设 $x\in\operatorname{bd}F$. 由于 F 是闭的, 有 $x\in F$, 因此 $\rho_F(x)\leqslant 1$. 相反地, 假设 $\gamma:=\rho_F(x)<1$, 则存在实数序列 $\{t_k\}$ 收敛于 γ 且对任意的 $k, x\in t_kF$. 于是 $x\in\gamma F=\gamma F+(1-\gamma)0\subset F$. 根据命题 3.32, Minkowski 度规 ρ_F 是连续的, 因此集合 $\{u\in\mathbb{R}^n|\rho_F(u)<1\}$ 是包含 x 的 F 的开子集. 因此 $x\in\operatorname{int}F$, 这就完成了证明. □

进一步, 回顾集合 $Q\subset\mathbb{R}^n$ 称为是严格凸的, 如果对任意的 $x,y\in Q$ 且 $x\neq y$, 对任意的 $t\in(0,1)$ 有 $tx+(1-t)y\in\operatorname{int}Q$.

命题 4.40 假设 F 是 $\mathbb{R}^n$ 的闭有界严格凸子集且满足 $0\in\operatorname{int}F$, 则

$$\rho_F(x+y)=\rho_F(x)+\rho_F(y), \quad x,y\neq 0 \tag{4.35}$$

当且仅当对某 $\lambda>0, x=\lambda y$.

证明 记 $\alpha:=\rho_F(x), \beta:=\rho_F(y)$, 注意到由于 F 是有界的, 所以 $\alpha,\beta>0$. 如果线性性质 (4.35) 成立, 则

$$\rho_F\left(\frac{x+y}{\alpha+\beta}\right)=1,$$

相应也蕴涵着

$$\rho_F\left(\frac{x}{\alpha}\frac{\alpha}{\alpha+\beta}+\frac{y}{\beta}\frac{\beta}{\alpha+\beta}\right)=1.$$

从而有

$$\frac{x}{\alpha}\frac{\alpha}{\alpha+\beta}+\frac{y}{\beta}\frac{\beta}{\alpha+\beta}\in \operatorname{bd}F.$$

由于 $\frac{x}{\alpha}\in F, \frac{y}{\beta}\in F$, 且 $\frac{\alpha}{\alpha+\beta}\in(0,1)$, 由 F 的严格凸性得

$$\frac{x}{\alpha}=\frac{y}{\beta}, \quad 即 \quad x=\lambda y \quad 且 \quad \lambda=\frac{\rho_F(x)}{\rho_F(y)}.$$

命题中反向的蕴涵关系是显然的. □

命题 4.41 设 F 与命题 4.40 的相同, 设 Ω 是 $\mathbb{R}^n$ 的非空闭凸子集, 则对任意的 $x\in\mathbb{R}^n$, 集合 $\Pi_F(x;\Omega)$ 是单点集.

证明 只需考虑 $x\notin\Omega$ 的情形. 显然 $\Pi_F(x;\Omega)\neq\varnothing$. 相反地假设存在 $u_1,u_2\in\Pi_F(x;\Omega)$ 且 $u_1\neq u_2$, 则

$$\mathcal{T}_F(x;\Omega)=\rho_F(u_1-x)=\rho_F(u_2-x)=r>0,$$

这蕴涵着 $\frac{u_1-x}{r}\in \operatorname{bd}F$ 和 $\frac{u_2-x}{r}\in\operatorname{bd}F$. 由于 F 是严格凸的, 有

$$\frac{1}{2}\frac{u_1-x}{r}+\frac{1}{2}\frac{u_2-x}{r}=\frac{1}{r}\left(\frac{u_1+u_2}{2}-x\right)\in\operatorname{int}F,$$

因此有 $\rho_F(u-x)<r=\mathcal{T}_F(x;\Omega)$ 且 $u:=\frac{u_1+u_2}{2}\in\Omega$, 矛盾, 这就完成了命题的证明. □

当集合 F 和 Ω 都是严格凸的时, 下面中以广义投影 $\Pi_F(x;\Omega)$ 表示其唯一元素.

命题 4.42 设 F 是 $\mathbb{R}^n$ 的闭有界凸子集且满足 $0\in\operatorname{int}F$, 设 $\Omega\subset\mathbb{R}^n$ 是非空闭集且 $\bar{x}\notin\Omega$. 如果 $\bar{\omega}\in\Pi_F(\bar{x};\Omega)$, 则 $\bar{\omega}\in\operatorname{bd}\Omega$.

证明 假设 $\bar{\omega}\in\operatorname{int}\Omega$, 则存在 $\varepsilon>0$ 使得 $\mathbb{B}(\bar{\omega};\varepsilon)\subset\Omega$. 定义

$$z:=\bar{\omega}+\varepsilon\frac{\bar{x}-\bar{\omega}}{\|\bar{x}-\bar{\omega}\|}\in\mathbb{B}(\bar{\omega};\varepsilon),$$

并且从 Minkowski 度规 (3.11) 的构造注意到

$$\rho_F(z-\bar{x})=\rho_F\left(\bar{\omega}+\varepsilon\frac{\bar{x}-\bar{\omega}}{\|\bar{x}-\bar{\omega}\|}-\bar{x}\right)=\left(1-\frac{\varepsilon}{\|\bar{x}-\bar{\omega}\|}\right)\rho_F(\bar{\omega}-\bar{x})<\rho_F(\bar{\omega}-\bar{x})=\mathcal{T}_F(\bar{x};\Omega).$$

这表明 $\bar{\omega}$ 必为 Ω 的边界点. □

命题 4.43 在问题 (4.34) 的情形下假设集合 Ω_i, $i=1,\cdots,m$ 是严格凸的. 如果对任意的 $x,y\in\Omega_0$ 且 $x\neq y$, 存在 $i_0\in\{1,\cdots,m\}$ 使得 $\mathcal{L}(x,y)$, 有

$$\mathcal{L}(x,y)\cap\Omega_{i_0}=\varnothing,$$

其中 $\mathcal{L}(x,y)$ 是连接 x 和 y 的直线, 则式 (4.34) 中定义的函数

$$\mathcal{H}(x)=\sum_{i=1}^{m}\mathcal{T}_F(x;\Omega_i)$$

在约束集合 Ω_0 上是严格凸的 (图 4.3).

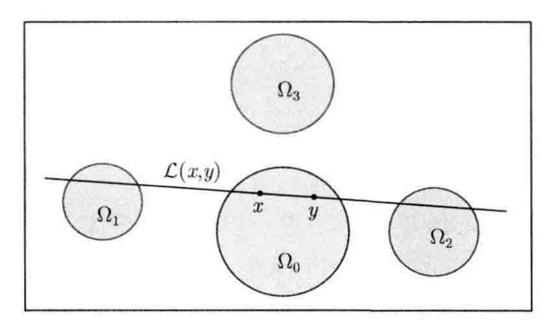


图 4.3 满足命题 4.43 假设的三个集合

证明 相反地假设存在 $x,y\in\Omega_0, x\neq y, t\in(0,1)$ 使得

$$\mathcal{H}(tx+(1-t)y)=t\mathcal{H}(x)+(1-t)\mathcal{H}(y).$$

由每个函数 $\mathcal{T}_F(x;\Omega_i), i=1,\cdots,m$ 的凸性, 有

$$\mathcal{T}_F(tx+(1-t)y;\Omega_i)=t\mathcal{T}_F(x;\Omega_i)+(1-t)\mathcal{T}_F(y;\Omega_i),\quad i=1,\cdots,m. \tag{4.36}$$

进一步, 对某 $i_0\in\{1,\cdots,m\}$, 假设 $\mathcal{L}(x,y)\cap\Omega_{i_0}=\varnothing$, 并定义

$$u:=\Pi_F(x;\Omega_{i_0}),\quad v:=\Pi_F(y;\Omega_{i_0}).$$

则由式 (4.36)、定理 3.33 和 Minkowski 度规的凸性得

$$\begin{aligned}t\rho_F(u-x)+(1-t)\rho_F(v-y)&=t\mathcal{T}_F(x;\Omega_{i_0})+(1-t)\mathcal{T}_F(y;\Omega_{i_0})\\&=\mathcal{T}_F(tx+(1-t)y;\Omega_{i_0})\\&\leqslant\rho_F(tu+(1-t)v-(tx+(1-t)y))\\&\leqslant t\rho_F(u-x)+(1-t)\rho_F(v-y).\end{aligned}$$

这表明 $tu+(1-t)v=\Pi_F(tx+(1-t)y;\Omega_{i_0})$. 因此 $u=v$, 由于否则有 $tu+(1-t)v\in\operatorname{int}\Omega_{i_0}$, 矛盾. 从而 $u=v=\Pi_F(tx+(1-t)y;\Omega_{i_0})$. 再次利用式 (4.36) 得

$$\rho_F(u-(tx+(1-t)y))=\rho_F(t(u-x))+\rho_F((1-t)(u-y)),$$

由于 $x, y \notin \Omega_{i_0}$, 所以 $u-x, u-y \neq 0$. 根据命题 4.40, 找到 $\lambda > 0$ 使得 $t(u-x) = \lambda(1-t)(u-y)$, 从而

$$u-x=\gamma(u-y) \text{ 且 } v=\frac{\lambda(1-t)}{t}\neq 1.$$

这就证明了包含关系

$$u=\frac{1}{1-\gamma}x-\frac{\gamma}{1-\gamma}y\in\mathcal{L}(x,y)$$

成立, 这是一个矛盾, 从而完成了命题的证明. □

下面将建立问题 (4.34) 的存在与唯一性定理.

命题 4.44 在命题 4.43 的情形下进一步假设集合 $\Omega_i, i=0,\cdots,m$ 中至少有一个是有界的, 则广义 Fermat-Torricelli 问题 (4.34) 有唯一解.

证明 根据命题 4.43 中 $\mathcal{H}$ 的严格凸性, 只需证明问题 (4.34) 有最优解. 如果 Ω_0 是有界的, 则它是紧的, 又由于根据推论 3.34 问题 (4.34) 中的函数 $\mathcal{H}$ 是 Lipschitz 连续的, 结论由 (Weierstrass) 定理 4.10 可得. 现考虑 $\Omega_i, i=1,\cdots,m$ 中有一个集合是有界的情形; 假设它是 Ω_1. 此时对任意的 $\alpha\in\mathbb{R}$, 集合

$$\mathcal{L}_\alpha=\{x\in\Omega_0\mid\mathcal{H}(x)\leqslant\alpha\}\subset\{x\in\Omega_0\mid\mathcal{T}_F(x;\Omega_1)\leqslant\alpha\}\subset\Omega_0\cap(\Omega_1-\alpha F)$$

是紧的. 根据定理 4.10 这就保证问题 (4.34) 的极小值点的存在性. □

下面借助于凸分析和最优化完整地解决下面问题 (4.34) 的简化版本:

$$\min\ \mathcal{H}(x)=\sum_{i=1}^{3}d(x;\Omega_i),\quad x\in\mathbb{R}^2, \tag{4.37}$$

其中目标集合 $\Omega_i := \mathbb{B}(c_i;r_i), i=1,2,3$ 是 $\mathbb{R}^2$ 的闭球, 为避免平凡性它们的中心 c_i 假定为互不相同的.

下面首先研究这个问题的性质以便根据初始数据 (c_i, r_i) 得到可验证的条件确保问题 (4.37) 解的唯一性.

命题 4.45 问题 (4.37) 的解集 S 是 $\mathbb{R}^2$ 中的非空紧凸集合.

证明 问题 (4.37) 解的存在性由定理 4.44 的证明可得. 还易见 S 是闭的且有界的. S 的凸性可由距离函数 $d(x;\Omega_i)$ 的凸性推得. □

接下来, 对任意的 $u\in\mathbb{R}^2$, 定义指标子集

$$\mathcal{I}(u):=\{i\in\{1,2,3\}|u\in\Omega_i\}, \tag{4.38}$$

这在下面将被广泛应用.

命题 4.46 假设 $\bar{x}$ 不属于任何集合 $\Omega_i, i=1,2,3$, 即 $\mathcal{I}(\bar{x})=\varnothing$, 则 $\bar{x}$ 是问题 (4.37) 的最优解当且仅当 $\bar{x}$ 是由式 (4.22) 中以 $a_i := c_i(i=1,2,3)$ 为中心生成的经典 Fermat-Torricelli 问题 (4.23) 的解.

证明 假设 $\bar{x}$ 是问题 (4.37) 的解且 $\bar{x} \notin \Omega_i, i=1,2,3$. 选取 $\delta > 0$ 使得当 $i=1,2,3$ 时 $\mathbb{B}(\bar{x};\delta) \cap \Omega_i = \varnothing$, 则对任意的 $x \in \mathbb{B}(\bar{x};\delta)$ 可得到下式:

$$\sum_{i=1}^{3}\|\bar{x}-c_i\| - \sum_{i=1}^{3} r_i = \inf_{x\in\mathbb{R}^2}\mathcal{H}(x) = \mathcal{H}(\bar{x}) \leqslant \mathcal{H}(x) = \sum_{i=1}^{3}\|x-c_i\| - \sum_{i=1}^{3} r_i.$$

这表明 $\bar{x}$ 是问题

$$\min \mathcal{F}(x) := \sum_{i=1}^{3}\|x-c_i\|, \quad x \in \mathbb{R}^2 \tag{4.39}$$

的局部极小值点. 由 F 的凸性得它也是该问题的全局极小值点. 从而 $\bar{x}$ 是由 c_1, c_2, c_3 生成的经典 Fermat-Torricelli 问题 (4.23) 的解. 相反的蕴涵关系也是易证的. □

利用式 (4.38) 定义的指标子集给出问题 (4.37) 解唯一性的条件.

命题 4.47 设 $\bar{x}$ 是问题 (4.37) 的解使得式 (4.38) 中的 $\mathcal{I}(\bar{x})=\varnothing$, 则 $\bar{x}$ 是问题 (4.37) 的唯一解, 即 $\xi=\{\bar{x}\}$.

证明 相反地假设存在 $u \in \xi$ 使得 $u \neq \bar{x}$. 由于 $\mathcal{I}(\bar{x})=\varnothing$, 对 $i=1,2,3$ 有 $\bar{x} \notin \Omega_i$. 则由命题 4.46 知 $\bar{x}$ 是由 c_1, c_2, c_3 生成的经典 Fermat-Torricelli 问题的解. 根据命题 4.45 中 ξ 的凸性得 $[\bar{x}, u] \subset \xi$. 于是可找到 $v \neq \bar{x}$ 使得 $v \in \xi$ 且 $v \notin \Omega_i (i=1,2,3)$. 这表明 v 也是经典 Fermat-Torricelli 问题 (4.39) 的解, 与问题 (4.39) 的唯一性结果矛盾, 从而证明了命题的断言是正确的. □

下面的命题证明椭球体的一个直观的几何性质.

命题 4.48 给定 $\alpha > 0, a, b \in \mathbb{R}^2$ 且 $a \neq b$, 考虑集合

$$E := \{x \in \mathbb{R}^2 \mid \|x-a\| + \|x-b\| = \alpha\},$$

并且假设 $\|a-b\| < \alpha$, 则对 $x, y \in E$ 且 $x \neq y$, 有

$$\left\|\frac{x+y}{2} - a\right\| + \left\|\frac{x+y}{2} - b\right\| < \alpha,$$

特别地, 这保证 $\dfrac{x+y}{2} \notin E$.

证明 由三角不等式得

$$\begin{aligned}\left\|\frac{x+y}{2} - a\right\| + \left\|\frac{x+y}{2} - b\right\| &= \frac{1}{2}(\|x-a+y-a\| + \|x-b+y-b\|)\\ &\leqslant \frac{1}{2}(\|x-a\| + \|y-a\| + \|x-b\| + \|y-b\|) = \alpha.\end{aligned}$$

用反证法, 假设

$$\left\|\frac{x+y}{2}-a\right\|+\left\|\frac{x+y}{2}-b\right\|=\alpha, \tag{4.40}$$

即 $\frac{x+y}{2}\in E$. 由于范数 (总) 是 Euclid 范数, 由于 $x\neq y$, 存在常数 $\beta,\gamma\in(0,\infty)\backslash\{1\}$ 使得

$$x-a=\beta(y-a), \quad x-b=\gamma(y-b).$$

这相应地蕴涵着

$$a=\frac{1}{1-\beta}x-\frac{\beta}{1-\beta}y\in\mathcal{L}(x,y), \quad b=\frac{1}{1-\gamma}x-\frac{\gamma}{1-\gamma}y\in\mathcal{L}(x,y).$$

由于 $\alpha>\|a-b\|$, 可看到 $x,y\in\mathcal{L}(a,b)\backslash[a,b]$. 现验证 $\frac{x+y}{2}\in[a,b]$, 若这不成立, 则不失一般性, 可设这些点依这样的顺序排列: $x,\frac{x+y}{2},a,b$ 与 y, 则

$$\left\|\frac{x+y}{2}-a\right\|+\left\|\frac{x+y}{2}-b\right\|<\|x-a\|+\|x-b\|=\alpha$$

与式 (4.40) 矛盾. 因此 $\frac{x+y}{2}\in[a,b]$, 但此时有

$$\left\|\frac{x+y}{2}-a\right\|+\left\|\frac{x+y}{2}-b\right\|=\|a-b\|<\alpha,$$

这是一个矛盾. 于是 $\frac{x+y}{2}\notin E$, 从而完成了证明. □

接下来的结果利用指标集 (4.38) 也确保问题 (4.37) 解的唯一性, 它不同于命题 4.47.

命题 4.49　设 S 是问题 (4.37) 的解集, 设 $\bar{x}\in S$ 满足条件: $\mathcal{I}(\bar{x})=\{i\}$ 且 $\bar{x}\notin[c_j,c_k]$, 其中 i,j,k 是 $\{1, 2, 3\}$中互不相同的指标, 则 S 是单点集, 即 $S=\{\bar{x}\}$.

证明　为明确起见, 假设 $\mathcal{I}(\bar{x})=\{1\}$, 则有 $\bar{x}\in\Omega_1,\bar{x}\notin\Omega_2,\bar{x}\notin\Omega_3$, 并且 $\bar{x}\notin[c_2,c_3]$. 记

$$\alpha:=\|\bar{x}-c_2\|+\|\bar{x}-c_3\|=\inf_{x\in\mathbb{R}^2}\mathcal{H}(x)+r_2+r_3,$$

并注意到 $\alpha>\|c_2-c_3\|$. 考虑椭圆

$$E:=\{x\in\mathbb{R}^2|\|x-c_2\|+\|x-c_3\|=\alpha\},$$

则得 $\bar{x}\in E\cap\Omega_1$. 用反证法证明, 假设 S 不是单点集, 即存在 $u\in S$, 满足 $u\neq\bar{x}$. ξ 的凸性蕴涵着 $[\bar{x},u]\subset S$. 可选取充分接近于 $\bar{x}$ 但不同于它的 $v\in[\bar{x},u]$ 使得 $[\bar{x},v]\cap\Omega_2=\varnothing$ 且 $[\bar{x},v]\cap\Omega_3=\varnothing$. 下面先证 $v\notin\Omega_1$. 假设相反 $d(v;\Omega_1)=0$, 于是

$$\mathcal{H}(v)=\inf_{x\in\mathbb{R}^2}\mathcal{H}(x)=d(v;\Omega_2)+d(v;\Omega_3)=\|v-c_2\|-\gamma_2+\|v-c_3\|-\gamma_3,$$

这蕴涵着 $v \in E$. 由 S 与 Ω_1 的凸性得 $\dfrac{\bar{x}+v}{2} \in S \cap \Omega_1$. 进一步有 $\dfrac{\bar{x}+v}{2} \notin \Omega_2$ 且 $\dfrac{\bar{x}+v}{2} \notin \Omega_3$, 从而 $\dfrac{\bar{x}+v}{2} \in E$. 但是根据命题 4.48 这是不可能发生的. 因此, $v \in S$ 且 $\mathcal{I}(v) = \varnothing$, 与命题 4.47 的结果矛盾, 这就证明了 S 是单点集. □

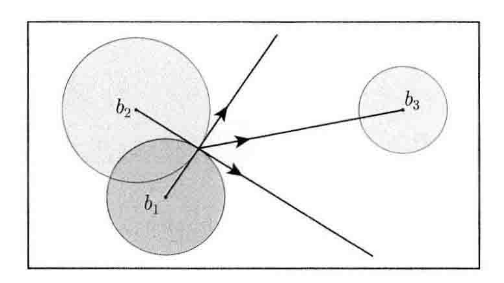


图 4.4 $|\mathcal{I}(\bar{x})| = 2$ 时的 Fermat-Torricelli 问题

下证所期望的问题 (4.37) 最优解的边界性性质.

命题 4.50 设 $\bar{x} \in S$ 且设 $\mathcal{I}(\bar{x}) = \{i, j\} \subset \{1, 2, 3\}$, 则 $\bar{x} \in \text{bd}\,(\Omega_i \cap \Omega_j)$.

证明 为明确起见假设 $\mathcal{I}(\bar{x}) = \{1, 2\}$, 即 $\bar{x} \in \Omega_1 \cap \Omega_2$ 且 $\bar{x} \notin \Omega_3$. 利用反证法, 假设 $\bar{x} \in \text{int}\,(\Omega_1 \cap \Omega_2)$, 因此存在 $\delta > 0$ 使得 $\mathbb{B}(\bar{x}; \delta) \subset \Omega_1 \cap \Omega_2$. 记 $p := \Pi(\bar{x}; \Omega_3), \gamma := \|\bar{x} - p\| > 0$, 且取 $q \in [\bar{x}, p] \cap \mathbb{B}(\bar{x}; \delta)$ 满足条件 $\|q - p\| < \|\bar{x} - p\| = d(\bar{x}; \Omega_3)$, 则得

$$\mathcal{H}(q) = \sum_{i=1}^{3} d(q; \Omega_i) = d(q; \Omega_3) \leqslant \|q - p\| < \|\bar{x} - p\| = d(\bar{x}; \Omega_3) = \sum_{i=1}^{3} d(\bar{x}; \Omega_i) = \mathcal{H}(\bar{x}),$$

与 $\bar{x} \in S$ 矛盾, 从而完成了证明. □

下面的结果给出问题 (4.37) 解的唯一性的可验证的必要和充分条件.

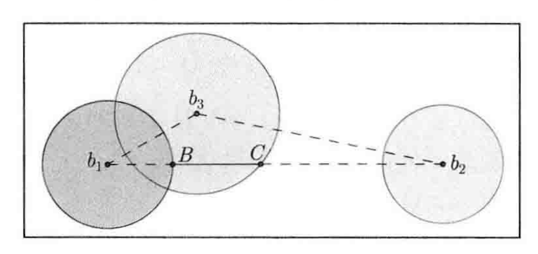


图 4.5 具有无穷多个解的广义 Fermat-Torrilelli 问题

定理 4.51 设 $\bigcap_{i=1}^{3} \Omega_i = \varnothing$, 则问题 (4.37) 有不止一个解当且仅当对互不相同的指标 $i, j, k \in \{1, 2, 3\}$, 存在一个集合 $[c_i, c_j] \cap \Omega_k$ 包含属于 Ω_k 内部的点 $u \in \mathbb{R}^2$, 且使得指标集 $\mathcal{I}(u)$ 是单点集.

证明 设 $u \in [c_1, c_2] \cap \Omega_3, u \in \text{int}\,\Omega_3$ 满足定理"当"部分的条件, 则 $u \notin \Omega_1, u \notin \Omega_2$, 且可选取 $v \neq u$ 满足 $v \in [c_1, c_2], v \notin \Omega_1, v \notin \Omega_2, v \in \text{int}\,\Omega_3$. 这表明集合 $\mathcal{I}(v)$ 是

单点集. 下证此时 u 是问题 (4.37) 的解. 为此, 首先注意到

$$\mathcal{H}(u) = d(u;\Omega_1) + d(u;\Omega_2) + d(u;\Omega_3) = \|c_1 - c_2\| - r_1 - r_2.$$

选取 $\delta > 0$ 满足 $\mathbb{B}(u;\delta) \cap \Omega_1 = \varnothing, \mathbb{B}(u;\delta) \cap \Omega_2 = \varnothing$ 且 $\mathbb{B}(u;\delta) \subset \Omega_3$. 对任意的 $x \in \mathbb{B}(u;\delta)$, 得

$$\mathcal{H}(x) = d(x;\Omega_2) + d(x;\Omega_3) = \|x - c_1\| + \|x - c_2\| - r_1 - r_2 \geqslant \|c_1 - c_2\| - r_1 - r_2 = \mathcal{H}(u),$$

这蕴涵着 u 是 $\mathcal{H}$ 的局部 (因此全局) 最优解. 同理可证 $v \in S$, 因此解集 S 不是单点集.

为证"仅当"部分, 假设 S 不是单点集, 选取两个不同的元素 $\bar{x}, \bar{y} \in S$, 根据命题 4.45 有 $[\bar{x}, \bar{y}] \subset S$. 如果存在式 (4.37) 的一个解不属于任何一个 Ω_i, $i \in \{1,2,3\}$. 则根据命题 4.47, S 简化为单点集, 矛盾. 于是确定起见可假设 Ω_1 包含无穷多个解, 则根据球 Ω_1 明显的严格凸性它的内部 $\operatorname{int}\Omega_1$ 也包含无穷多个解. 如果有这样的一个解 u 使得 $\mathcal{I}(u) = \{1\}$, 则 $u \in \operatorname{int}\Omega_1, u \notin \Omega_2$ 且 $u \notin \Omega_3$. 这蕴涵着 $u \in [c_2, c_3]$, 这是因为根据命题 4.49 相反的结论保证解的唯一性. 考虑对任意属于 $\operatorname{int}\Omega_1$ 的解集合 $\mathcal{I}(u)$ 都包含两个元素的情形, 则存在属于这两个集合交的无穷多个解, 因为这两个集合的交是严格凸的. 所以存在问题 (4.37) 的一个解属于这个交的内部, 这与命题 4.50 矛盾, 从而完成了定理的证明. □

现考虑问题 (4.34) 的另一个特殊情形, 其中的常数动态 F 由闭单位球给出. 该问题可描述为

$$\min \mathcal{D}(x) := \sum_{i=1}^{m} d(x;\Omega_i), \quad \text{s.t.} \quad x \in \Omega_0. \tag{4.41}$$

对 $x \in \mathbb{R}^n$, 定义 $\{1, \cdots, m\}$ 的指标子集为

$$I(x) := \{i \in \{1, \cdots, m\} \mid x \in \Omega_i\}, \quad J(x) := \{i \in \{1, \cdots, m\} \mid x \notin \Omega_i\}. \tag{4.42}$$

显然对任意的 $x \in \mathbb{R}^n, I(x) \cup J(x) = \{1, \cdots, m\}$ 且 $I(x) \cap J(x) = \varnothing$. 下面的定理利用到 Ω_i 的投影完全地刻画了问题 (4.41) 的最优解.

定理 4.52 对问题 (4.41) 下面的断言成立:

(i) 设 $\bar{x} \in \Omega_0$ 是问题 (4.41) 的解. 若 $i \in J(\bar{x})$ 时, 记 $a_i(\bar{x}) \equiv A_i(\bar{x})$, 则有

$$-\sum_{i \in J(\bar{x})} a_i(\bar{x}) \in \sum_{i \in I(\bar{x})} A_i(\bar{x}) + N(\bar{x};\Omega_0), \tag{4.43}$$

其中集合 $A_i(\bar{x})$, $i = 1, \cdots, m$ 通过

$$A_i(\bar{x}) := \begin{cases} \dfrac{\bar{x} - \Pi(\bar{x};\Omega_i)}{d(\bar{x};\Omega_i)}, & \bar{x} \notin \Omega_i, \\ N(\bar{x};\Omega_i) \cap \mathbb{B}, & \bar{x} \in \Omega_i \end{cases} \tag{4.44}$$

来计算.

(ii) 相反地, 如果 $\bar{x} \in \Omega_0$ 满足式 (4.43), 则 $\bar{x}$ 是问题 (4.41) 的解.

证明 为证 (i) 成立, 取定问题 (4.41) 的一个最优解 $\bar{x}$ 并等价地描述它为无约束最优化问题:

$$\min \mathcal{D}(x) + \delta(x; \Omega), \quad x \in \mathbb{R}^n \tag{4.45}$$

的一个最优解. 对问题 (4.45) 应用次微分 Fermat 法则, $\bar{x}$ 可通过

$$0 \in \partial \left(\sum_{i=1}^{n} d(\cdot; \Omega_i) + \delta(\cdot; \Omega) \right) (\bar{x}) \tag{4.46}$$

刻画. 由于所有的函数 $d(\cdot; \Omega_i), i = 1, \cdots, n$ 是凸连续的, 对式 (4.46) 利用定理 2.15 的次微分和法则, 得

$$\begin{aligned}
0 &\in \sum_{i \in J(\bar{x})} \partial d(\bar{x}; \Omega_i) + \sum_{i \in I(\bar{x})} d(\bar{x}; \Omega_i) + N(\bar{x}; \Omega_0) \\
&= \sum_{i \in J(\bar{x})} A_i(\bar{x}) + \sum_{i \in I(\bar{x})} A_i(\bar{x}) + N(\bar{x}; \Omega_0) \\
&= \sum_{i \in J(\bar{x})} a_i(\bar{x}) + \sum_{i \in I(\bar{x})} A_i(\bar{x}) + N(\bar{x}; \Omega_0),
\end{aligned}$$

这就证明了所断言的包含关系

$$-\sum_{i \in J(\bar{x})} a_i(\bar{x}) \in \sum_{i \in I(\bar{x})} A_i(\bar{x}) + N(\bar{x}; \Omega_0)$$

成立.

相反的断言 (ii) 通过考虑 (i) 的证明中每一步的“当且仅当”性质类似可证. □

下面特别地考虑问题 (4.41) 中的约束集 Ω 与目标集 Ω_i 不相交的情形.

推论 4.53 考虑问题 (4.41), 其中 $\Omega_0 \cap \Omega_i = \varnothing, i = 1, \cdots, m$. 给定 $\bar{x} \in \Omega_0$, 定义

$$a_i(\bar{x}) := \frac{\bar{x} - \Pi(\bar{x}; \Omega_i)}{d(\bar{x}; \Omega_i)} \neq 0, \quad i = 1, \cdots, m, \tag{4.47}$$

则 $\bar{x} \in \Omega_0$ 是问题 (4.41) 的最优解当且仅当

$$-\sum_{i=1}^{m} a_i(\bar{x}) \in N(\bar{x}; \Omega_0). \tag{4.48}$$

证明 此时有 $\bar{x} \notin \Omega_i, i = 1, \cdots, m$, 这意味着 $I(\bar{x}) = \varnothing$ 且 $J(\bar{x}) = \{1, \cdots, m\}$. 于是我们的断言是定理 4.52 的直接结果. □

对问题 (4.41) 解的更多细节, 考虑当 $m = 3$ 时的特殊推论.

定理 4.54　在定理 4.52 的框架下, 设 $m=3$, 这里目标集合 Ω_1, Ω_2 和 Ω_3 是两两不交的, 且 $\Omega_0=\mathbb{R}^n$. 设 $a_i(\bar{x})$ 由式 (4.47) 定义, 则对问题 (4.41) 的最优解 $\bar{x}$, 下面的结论成立.

(i) 如果 $\bar{x}$ 属于集合 Ω_i 中的一个, 假设为 Ω_1, 则有

$$\langle a_2, a_3\rangle \leqslant -1/2 \text{ 且 } -a_2-a_3 \in N(\bar{x};\Omega_1).$$

(ii) 如果 $\bar{x}$ 不属于三个集合 Ω_1, Ω_2 与 Ω_3 中的任何一个, 则

$$\langle a_i, a_j\rangle = -1/2, \quad i \neq j, \quad i, j \in \{1,2,3\}.$$

反之, 如果 $\bar{x}$ 满足 (i) 或 (ii), 则它是问题 (4.41) 的最优解.

证明　为证 (i) 成立, 由定理 4.52 得 $\bar{x}\in\Omega_1$ 是问题 (4.41) 的解当且仅当

$$\begin{aligned}0 &\in \partial d(\bar{x};\Omega_1)+\partial d(\bar{x};\Omega_2)+\partial d(\bar{x};\Omega_3)\\ &= \partial d(\bar{x};\Omega_1)+a_2+a_3 = N(\bar{x};\Omega_1)\cap\mathbb{B}+a_2+a_3.\end{aligned}$$

于是 $-a_2-a_3\in N(\bar{x};\Omega_1)\cap\mathbb{B}$, 这等价于 $-a_2-a_3\in\mathbb{B}$ 且 $-a_2-a_3\in N(\bar{x};\Omega_1)$ 成立. 根据引理 4.31, 这也等价于

$$\langle a_2, a_3\rangle \leqslant -1/2 \text{ 且 } -a_2-a_3 \in N(\bar{x};\Omega_1),$$

从而 (i) 成立. 下面证 (ii), 当 $\bar{x}\notin\Omega_i, i=1,2,3$ 时类似地可得, 这个元素是问题 (4.41) 的解当且仅当

$$0\in\partial d(\bar{x};\Omega_1)+\partial d(\bar{x};\Omega_2)+\partial d(\bar{x};\Omega_3)=a_1+a_2+a_3,$$

则由引理 4.32 知上式等价于

$$\langle a_i, a_j\rangle = -1/2; \quad i\neq j, i, j\in\{1,2,3\}.$$

反之, (i) 或 (ii) 成立及 Ω_i 的凸性保证

$$0\in\partial d(\bar{x};\Omega_1)+\partial d(\bar{x};\Omega_2)+\partial d(\bar{x};\Omega_3)=\partial\mathcal{D}(\bar{x}),$$

从而 $\bar{x}$ 是所考虑问题的最优解. □

例 4.55　设问题 (4.37) 中的集合 Ω_1, Ω_2 与 Ω_3 是 $\mathbb{R}^n$ 中的闭球, 中心分别在 $(0,2), (-2,0)$ 与 $(2,0)$, 半径均为 $r=1$. 易见点 $(0,1)\in\Omega_1$ 满足定理 4.54(i) 中的所有条件, 因此它是该问题的最优解 (事实上是唯一解).

更一般地, 考虑 $\mathbb{R}^2$ 中的问题 (4.37), 由三个两两不交的圆盘 $\Omega_i, i=1,2,3$ 生成. 设 c_1, c_2 和 c_3 是这三个圆盘的中心. 先假设或者线段 $[c_2, c_3]\subset\mathbb{R}^2$ 与 Ω_1 相交, 或者 $[c_1, c_2]$ 与 Ω_2 相交, 或者 $[c_1, c_2]$ 与 Ω_3 相交. 不难验证交集中的任何点 (为明

确起见, 假设为集合 Ω_1 与 $[c_2, c_3]$ 的交) 是所考虑问题的最优解, 由于它满足定理 4.54(i) 的必要和充分最优性条件. 事实上, 如果 $\bar{x}$ 是这样的点, 则 a_2 和 a_3 是单位向量且 $\langle a_2, a_3\rangle = -1, -a_2 - a_3 = 0 \in N(\bar{x}; \Omega_1)$. 如果上面的交假设不成立, 可如下定义三个点 q_1, q_2 和 q_3. 设 u 和 v 分别是 $[c_1, c_2]$ 和 $[c_1, c_3]$ 与中心在 c_1 的圆盘边界的交点, 则存在由 u 和 v 决定的劣弧上的唯一点 q_1, 使得 $\angle c_1q_1c_2$ 与 $\angle c_1q_1c_3$ 相等. 点 q_2 和 q_3 可类似地定义. 由定理 4.54 得如果 $\angle c_2q_1c_3$ 或 $\angle c_1q_2c_3$ 或 $\angle c_2q_3c_1$ 等于或大于 $120°$(确定起见, 假设 $\angle c_2q_1c_3$ 是这样的角), 则点 $\bar{x} := q_1$ 是所考虑问题的最优解. 事实上, 此时式 (4.47) 中的 a_2 和 a_3 是单位向量, 满足 $\langle a_2, a_3\rangle \leqslant -1/2$ 且 $-a_2 - a_3 \in N(\bar{x}; \Omega_1)$, 因为向量 $-a_2 - a_3$ 与 Ω_1 正交.

如果这些角都不等于或大于 $120°$, 则存在点 q 不属于 $\Omega_i, i = 1, 2, 3$ 使得 $\angle c_1qc_2 = \angle c_2qc_3 = \angle c_3qc_1 = 120°$, 而且 q 是所考虑问题的最优解. 此时注意到点 q 也是由点 c_1, c_2 和 c_3 确定的经典 Fermat-Torricelli 问题的唯一最优解; 参见图 4.6.

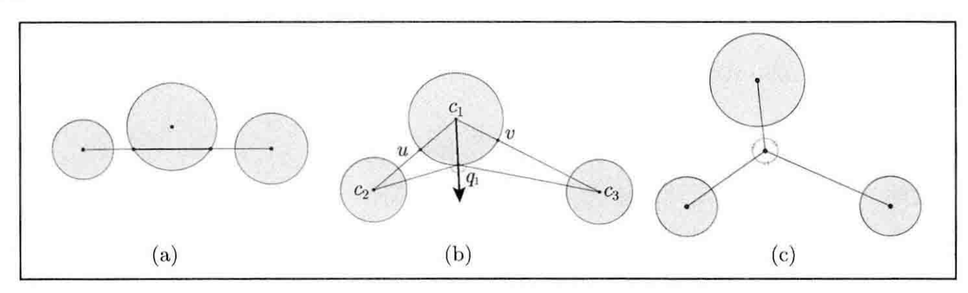


图 4.6　关于三个圆盘的广义 Fermat-Torricelli 问题

在本节最后, 讨论 4.3 节中次梯度算法在广义 Fermat-Torricelli 问题 (4.34) 数值解中的应用.

定理 4.56　考虑广义 Fermat-Torricelli 问题 (4.34), 假设 $\xi \neq \varnothing$ 是它的解集. 选取正数序列 $\{\alpha_k\}$ 和初始点 $x_1 \in \Omega_0$, 构造迭代序列

$$x_{k+1} := \Pi\left(x_k - \alpha_k \sum_{i=1}^{m} v_{ik}; \Omega_0\right), \quad k = 1, 2, \cdots, \tag{4.49}$$

其中, 如果 $x_k \notin \Omega_i$, 对某 $\omega_{ik} \in \Pi_F(x_k; \Omega_i)$, 任意选取次梯度元素

$$v_{ik} \in -\partial\rho_F(\omega_{ik} - x_k) \cap N(\omega_{ik}; \Omega_i), \tag{4.50}$$

否则, 令 $v_{ik} = 0$, 这里 $\Pi_F(x; \Omega)$ 表示广义投影算子 (3.21), ρ_F 是 Minkowski 度规 (3.11). 定义值序列

$$V_k := \min\{\mathcal{H}(x_j) \mid j = 1, \cdots, k\}.$$

(i) 如果序列 $\{\alpha_k\}$ 满足条件 (4.13), 则 $\{V_k\}$ 收敛于问题 (4.34) 的最优值 $\overline{V}$. 进

一步, 有估计

$$0 \leqslant V_k - \overline{V} \leqslant \frac{d(x_1;S)^2 + \ell^2 \sum_{k=1}^{\infty} \alpha_k^2}{2\sum_{i=1}^{k} \alpha_k},$$

这里 $\ell \geqslant 0$ 是费用函数 $\mathcal{H}(\cdot)$ 在 $\mathbb{R}^n$ 的 Lipschitz 常数.

(ii) 如果序列 $\{\alpha_k\}$ 满足条件 (4.15), 则 $\{V_k\}$ 收敛于最优值 $\overline{V}$, 且式 (4.49) 中的 $\{x_k\}$ 收敛于问题 (4.34) 的某最优解 $\bar{x} \in \xi$.

证明 这是命题 4.28 中所描述的投影次梯度算法在约束最优化问题 (4.34) 中的直接应用, 其中考虑到 4.3 节给出的凸最优化次梯度方法的结果. □

下面对问题 (4.34) 中 $F = \mathbb{B}$ 时的情形来具体说明上面的算法.

推论 4.57 在问题 (4.34) 中令 $F = \mathbb{B} \subset \mathbb{R}^n$, 设 $\{\alpha_k\}$ 是正数序列, $x_1 \in \Omega_0$ 是初始点. 考虑算法 (4.49), 其中 $\partial \rho_F(\cdot)$ 由

$$\partial \rho_F(\omega_{ki} - x_k) = \begin{cases} \dfrac{\Pi(x_k;\Omega_i) - x_k}{d(x_k;\Omega_i)}, & x_k \notin \Omega_i, \\ 0, & x_k \in \Omega_i \end{cases}$$

计算, 则对这种特殊情形定理 4.56 的所有结论成立.

证明 取 $\rho_F(x) = \|x\|$, 由定理 4.56 立即可得. □

下面举例说明定理 4.56 的次梯度算法在解由具有非球动态的极小时间函数描述的广义 Fermat-Torricelli 问题 (4.34) 中的应用. 值得注意的是, 为实施次梯度算法 (4.49), 仅需对每个目标集合计算式 (4.50) 右边集合中的一个次梯度. 为明确起见, 在下面考虑的例子中动态 F 由平面上的正方形 $[-1,1] \times [-1,1]$ 给出. 此时相应的 Minkowski 度规 (3.11) 由表达式

$$\rho_F(x_1,x_2) = \max\{|x_1|,|x_2|\}(x_1,x_2 \in \mathbb{R}^2) \tag{4.51}$$

给出.

例 4.58 考虑无约束广义 Fermat-Torricelli 问题 (4.34), 其中动态 $F = [-1,1] \times [-1,1]$, 右边位置的正方形目标 Ω_i(其中边平行于坐标轴) 以 (a_i, b_i) 为中心, r_i 为半径 (边的一半), $i = 1, \cdots, m$. 给定正数序列 $\{\alpha_k\}$ 和初始点 x_1, 根据算法 (4.49) 构造迭代 $x_k = (x_{1k}, x_{2k})$. 根据定理 3.39, 次梯度 v_{ik} 可如下选择:

$$v_{ik} = \begin{cases} (1,0), & |x_{2k} - b_i| \leqslant x_{1k} - a_i \text{ 且 } x_{1k} > a_i + r_i, \\ (-1,0), & |x_{2k} - b_i| \leqslant a_i - x_{1k} \text{ 且 } x_{1k} < a_i - r_i, \\ (0,1), & |x_{1k} - a_i| \leqslant x_{2k} - b_i \text{ 且 } x_{2k} > b_i + r_i, \\ (0,-1), & |x_{1k} - a_i| \leqslant b_i - x_{2k} \text{ 且 } x_{2k} < b_i - r_i, \\ (0,0), & \text{其他}. \end{cases}$$

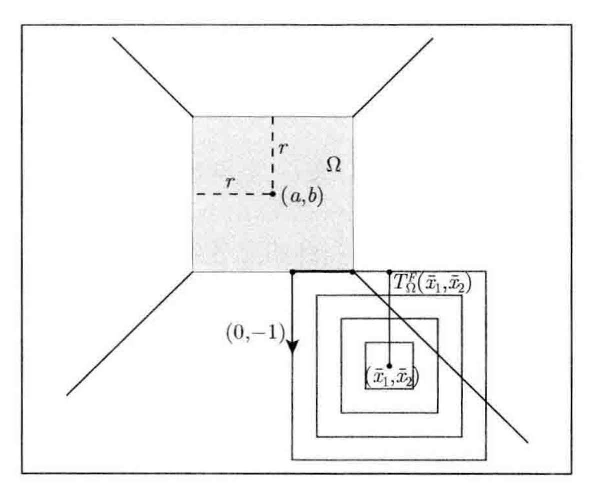


图 4.7 具有正方形动态的极小时间函数

4.6 广义 Sylvester 问题

在文献 [31] 中, James Joseph Sylvester 提出了最小封闭圆问题: 给定平面上有限多个点, 找到最小的圆围住所有的点. 一个多世纪以后, 这种类型的选址问题仍很活跃, 因为它们在数学上是漂亮的, 而且具有有意义的现实应用.

本节致力于利用凸分析致力于研究文献 [18] 中描述的下面的 Sylvester 问题的广义版本; 更进一步的发展也可参见文献 [19], [20]. 设 $F \subset \mathbb{R}^n$ 是包含原点作为内点的闭有界凸集. 定义中心在 $x \in \mathbb{R}^n$ 半径为 $r > 0$ 的广义球为

$$G_F(x;r) := x + rF.$$

显然, 对 $F = \mathbb{B}$, 广义球 $G_F(x;r)$ 减化为中点在 x 半径为 r 的闭球. 现给定有限多非空闭凸集合 Ω_i, $i = 0, \cdots, m$, 则广义 Sylvester 问题描述如下.

找到点 $x \in \Omega_0$ 和最小的数 $r \geqslant 0$ 使得

$$G_F(x;r) \cap \Omega_i \neq \varnothing, \quad \forall i = 1, \cdots, m.$$

注意到当 $\Omega_0 = \mathbb{R}^2$, $F = \mathbb{B}$ 且所有集合 $\Omega_i, i = 1, \cdots, m$ 是 $\mathbb{R}^2$ 中的单点集时, 广义 Sylvester 问题变为经典的 Sylvester 封闭圆问题. 集合 $\Omega_i, i = 1, \cdots, m$, 经常称为目标集合.

为研究广义 Sylvester 问题, 下面利用极小时间函数 (3.9) 引入费用函数

$$\mathcal{T}(x) := \max\{\mathcal{T}_F(x;\Omega_i) \mid i = 1, \cdots, m\},$$

并考虑下面的最优化问题:

$$\min \mathcal{T}(x), \quad \text{s.t.} \;\; x \in \Omega_0. \tag{4.52}$$

可证明广义球 $G_F(\bar{x};r)$, 其中 $\bar{x} \in \Omega_0$, 是广义 Sylvester 问题的最优解当且仅当 $\bar{x}$ 是最优化问题 (4.52) 的最优解, 其中 $r = \mathcal{T}(\bar{x})$.

本节命题给出了问题 (4.52) 解存在性的充分条件.

命题 4.59 广义 Sylvester 问题 (4.52) 有最优解, 如果对任意的 $\alpha \geqslant 0$, 集合 $\Omega_0 \cap [\bigcap_{i=1}^m (\Omega_i - \alpha F)]$ 是有界的. 特别地, 如果存在一个指标 $i_0 \in \{0, \cdots, m\}$ 使得 Ω_{i_0} 是紧的, 则该结论成立.

证明 对任意的 $\alpha \geqslant 0$, 考虑水平集

$$\mathcal{L}_\alpha := \{x \in \Omega_0 \mid \mathcal{T}(x) \leqslant \alpha\},$$

下面证明

$$\mathcal{L}_\alpha = \Omega_0 \cap \left[\bigcap_{i=1}^m (\Omega_i - \alpha F)\right]. \tag{4.53}$$

事实上, 对任意非空闭凸集合 $Q, \mathcal{T}_F(x;Q) \leqslant \alpha$, 当且仅当 $(x + \alpha F) \cap Q \neq \varnothing$, 这等价于 $x \in Q - \alpha F$. 如果 $x \in \mathcal{L}_\alpha$, 则 $x \in \Omega_0$ 且对 $i = 1, \cdots, m$, $\mathcal{T}_F(x;\Omega_i) \leqslant \alpha$. 因此 $x \in \Omega_i - \alpha F$, $i = 1, \cdots, m$, 从而证明了式 (4.53) 中包含关系“$\subset$”成立. 相反包含关系的证明也是直接的. 进一步, 显然集合 $\Omega_0 \cap [\bigcap_{i=1}^m (\Omega_i - \alpha F)]$ 对任意的 $\alpha \geqslant 0$ 的有界性保证关于式 (4.52) 中费用函数的水平集是有界的. 因此由定理 4.11 得问题 (4.52) 有最优解. 最后的断言是显然的. □

现在继续研究问题 (4.52) 最优解的唯一性, 下面把证明分为以下三个命题.

命题 4.60 考虑非空凸集 $\Omega \subset \mathbb{R}^n$ 上的凸函数 $g(x) \geqslant 0$, 则由 $h(x) := (g(x))^2$ 定义的函数在 Ω 上是严格凸的当且仅当 g 在任何线段 $[a,b] \subset \Omega$ 上不是常数, 其中 $a \neq b$.

证明 假设 h 在 Ω 上是严格凸的, 并假设相反 g 在线段 $[a,b] (a \neq b)$ 上是常数, 则 h 在这条线段上也是常数, 因此它不可能是严格凸的.

反之, 假设 g 在任何线段 $[a,b] \subset \Omega (a \neq b)$ 上不是常数. 注意到 h 是 Ω 上的凸函数, 因为它是在 $[0,\infty]$ 上不减的凸函数 $y = x^2$ 与值域是 $[0,\infty]$ 的子集的凸函数 g 的复合. 下证它在 Ω 上是严格凸的. 相反地假设存在 $t \in (0,1), x, y \in \Omega, x \neq y$, 满足

$$h(z) = th(x) + (1-t)h(y), \quad 其中 \quad z := tx + (1-t)y.$$

则 $(g(z))^2 = t(g(x))^2 + (1-t)(g(y))^2$. 由于 $g(z) \leqslant tg(x) + (1-t)g(y)$, 有

$$(g(z))^2 \leqslant t^2(g(x))^2 + (1-t)^2(g(y))^2 + 2t(1-t)g(x)g(y),$$

这显然蕴涵着

$$t(g(x))^2+(1-t)(g(y))^2 \leqslant t^2(g(x))^2+(1-t)^2(g(y))^2+2t(1-t)g(x)g(y).$$

因此 $(g(x)-g(y))^2 \leqslant 0$, 从而 $g(x)=g(y)$. 这表明对上述定义的点 $z \in (x,y)$, 有 $h(x)=h(z)=h(y)$. 下面证明 h 在线段 $[z,y]$ 上是常数得矛盾. 事实上, 任取 $u \in (z,y)$, 并注意到对某 $\nu \in (0,1)$, 有

$$h(u) \leqslant \nu h(z)+(1-\nu)h(y)=h(z).$$

另外, 由于 z 在 x 和 u 之间, 对某 $\mu \in (0,1)$, 有

$$h(z) \leqslant \mu h(x)+(1-\mu)h(u) \leqslant \mu h(z)+(1-\mu)h(z)=h(z),$$

因此 $h(z)=\mu h(x)+(1-\mu)h(u)=\mu h(z)+(1-\mu)h(u)$, 从而 $h(u)=h(z)$. 这与假设 $g(u)=\sqrt{h(u)}$ 在任何线段 $[a,b](a \neq b)$ 上不是常数矛盾, 从而完成了命题的证明.□

命题 4.61 设 $\mathbb{R}^n$ 中的集合 F 和 Ω 都是非空闭严格凸的, 其中 F 包含原点作为内点, 则凸函数 $\mathcal{T}_F(\cdot;\Omega)$ 在满足 $a \neq b$ 且 $[a,b] \cap \Omega = \varnothing$ 的任何线段 $[a,b] \subset \mathbb{R}^n$ 上不是常数.

证明 为证命题的断言成立, 相反地假设存在 $[a,b] \subset \mathbb{R}^n (a \neq b)$, $[a,b] \cap \Omega = \varnothing$ 且对任意的 $x \in [a,b]$ 满足 $\mathcal{T}_F(x;\Omega)=r$. 选取 $u \in \Pi_F(a;\Omega)$ 和 $v \in \Pi_F(b;\Omega)$, 得 $\rho_F(u-a)=\rho_F(v-b)=r$. 下证 $u \neq v$. 事实上, 若 $u=v$, 则有 $\rho_F(u-a)=\rho_F(u-b)=r$, 由 $r>0$ 可得

$$\rho_F\left(\frac{u-a}{r}\right)=\rho_F\left(\frac{u-b}{r}\right)=1,$$

于是 $\dfrac{u-a}{r} \in \operatorname{bd} F$, $\dfrac{u-b}{r} \in \operatorname{bd} F$. 由于 F 是严格凸的, 则有

$$\frac{1}{2}\left(\frac{u-a}{r}\right)+\frac{1}{2}\left(\frac{u-b}{r}\right)=\frac{1}{r}\left(u-\frac{a+b}{2}\right) \in \operatorname{int} F.$$

由于

$$\mathcal{T}_F\left(\frac{a+b}{2};\Omega\right) \leqslant \rho_F\left(u-\frac{a+b}{2}\right)<r,$$

这是一个矛盾. 为完成命题的证明, 取定 $t \in (0,1)$, 得

$$\begin{aligned} r=\mathcal{T}_F(ta+(1-t)b;\Omega) &\leqslant \rho_F(tu+(1-t)v-(ta+(1-t)b)) \\ &\leqslant t\rho_F(u-a)+(1-t)\rho_F(v-b)=r, \end{aligned}$$

这蕴涵着 $tu+(1-t)v \in \Pi_F(ta+(1-t)b;\Omega)$. 因此 $tu+(1-t)v \in \operatorname{bd}\Omega$, 与 Ω 的严格凸性矛盾, 这就完成了证明. □

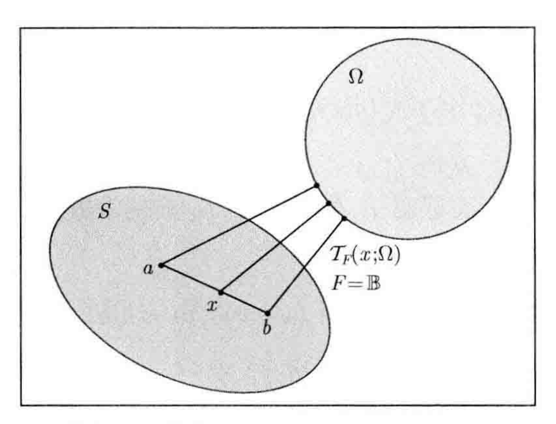


图 4.8 关于严格凸集的极小时间函数

命题 4.62 设 $\Omega \in \mathbb{R}^n$ 是非空闭凸集合, 设对 $i=1,\cdots,m$, $h_i:\Omega\to\mathbb{R}$ 是非负的连续凸函数. 定义

$$\phi(x):=\max\{h_1(x),\cdots,h_m(x)\},$$

并假设在某线段 $[a,b]\subset\Omega(a\neq b)$ 上 $\phi(x)=r>0$, 则存在线段 $[\alpha,\beta]\subset[a,b](\alpha\neq\beta)$ 和指标 $i_0\in\{1,\cdots,m\}$ 使得对任意的 $x\in[\alpha,\beta]$, 有 $h_{i_0}(x)=r$.

证明 当 $m=1$ 时结论自明. 当 $m=2$ 时, 考虑函数

$$\phi(x):=\max\{h_1(x),h_2(x)\}.$$

如果对任意的 $x\in[a,b], h_1(x)=r$, 则结论显然成立. 否则可找到 $x_0\in[a,b]$ 满足 $h_1(x_0)<r$ 和子区间 $[\alpha,\beta]\subset[a,b],\alpha\neq\beta$, 满足

$$h_1(x)<r,\quad \forall\, x\in[\alpha,\beta].$$

因此在这个子区间上 $h_2(x)=r$. 接下来利用数学归纳法, 假设结论对正整数 $m\geqslant 2$ 成立, 下证对 $m+1$ 个函数结论也成立.

记

$$\phi(x):=\max\{h_1(x),\cdots,h_m(x),h_{m+1}(x)\},$$

则有 $\phi(x)=\max\{h_1(x),k_1(x)\}$, 其中 $k_1(x):=\max\{h_2(x),\cdots,h_{m+1}(x)\}$, 而且由两个函数的情形和归纳法假设可得结论. □

现在已经准备好研究广义 Sylvester 问题 (4.52) 最优解的唯一性.

定理 4.63 在问题 (4.52) 的情形下, 进一步假设集合 F 和 $\Omega_i(i=1,\cdots,m)$ 是严格凸的, 则问题 (4.52) 至多有一个最优解当且仅当 $\bigcap_{i=0}^m\Omega_i$ 至多包含一个元素.

证明 定义 $\mathcal{K}(x) := (\mathcal{T}(x))^2$, 并考虑最优化问题

$$\min \mathcal{K}(x), \quad \text{s.t.} \quad x \in \Omega_0. \tag{4.54}$$

显然 $\bar{x} \in \Omega_0$ 是问题 (4.52) 的最优解当且仅当它是问题 (4.54) 的最优解. 下面将证明 $\bigcap_{i=0}^m \Omega_i$ 至多包含一个元素, 则 $\mathcal{K}$ 在 Ω_0 上是严格凸的, 从而问题 (4.52) 至多有一个最优解. 事实上，假设 $\mathcal{K}$ 在 Ω_0 上不是严格凸的. 根据命题 4.60, 可找到线段 $[a,b] \subset \Omega_0 (a \neq b)$ 和数 $r \geqslant 0$ 满足

$$\mathcal{T}(x) = \max\{\mathcal{T}_F(x;\Omega_i) \mid i = 1, \cdots, m\} = r, \quad \forall x \in [a,b].$$

显然 $r > 0$, 否则由于 $[a,b] \subset \bigcap_{i=0}^m \Omega_i$ 与这个集合至多包含一个元素的假设矛盾. 由命题 4.62 知存在线段 $[\alpha,\beta] \subset [a,b](\alpha \neq \beta)$ 和指标 $i_0 \in \{1,\cdots,m\}$, 使得

$$\mathcal{T}_F(x;\Omega_{i_0}) = r, \quad \forall x \in [\alpha,\beta].$$

由于 $r > 0$, 对任意的 $x \in [\alpha,\beta]$ 有 $x \notin \Omega_{i_0}$, 与命题 4.61 矛盾.

反之, 设问题 (4.52) 至多有一个解. 现假设集合 $\bigcap_{i=0}^m \Omega_i$ 包含不止一个元素. 显然 $\bigcap_{i=0}^m \Omega_i$ 是问题 (4.52) 的解集且最优值等于零. 因此这个问题有不止一个解, 矛盾, 这就完成了定理的证明. □

下面的推论给出广义 Sylvester 问题 (4.52) 最优解唯一性的一个充分条件.

推论 4.64 在问题 (4.52) 的情形下假设集合 F 和 $\Omega_i (i = 1, \cdots, m)$ 是严格凸的. 如果 $\bigcap_{i=0}^m \Omega_i$ 至多包含一个元素, 而且如果 $\Omega_i, i = 0, \cdots, m$ 中有一个集合是有界的, 则问题 (4.52) 有唯一的最优解.

证明 由于 $\bigcap_{i=0}^m \Omega_i$ 至多包含一个元素，根据定理 4.63 问题 (4.52) 至多有一个最优解. 因为根据命题 4.59 这个解集是非空的, 所以刚好有唯一解. □

下面考虑对非空闭凸集 Ω_i, 在假设 $\bigcap_{i=1}^n \Omega_i = \varnothing$ 下由

$$\min \mathcal{M}(x) := \max\{d(x;\Omega_i) \mid i = 1, \cdots, m\}, \quad \text{s.t.} \quad x \in \mathbb{R}^n \tag{4.55}$$

给出问题 (4.52) 的一个特殊无约束情形. 当 $F = \mathbb{B} \subset \mathbb{R}^n$ 时, 问题 (4.55) 等价于问题 (4.52). 记问题 (4.55) 的活跃指标集为

$$I(x) := \{i \in \{1, \cdots, m\} \mid \mathcal{M}(x) = d(x;\Omega_i)\}, \quad x \in \mathbb{R}^n.$$

进一步, 对 $\forall x \notin \Omega_i$, 定义单点

$$a_i(x) := \frac{x - \Pi(x;\Omega_i)}{d(x;\Omega_i)}, \tag{4.56}$$

这里投影 $\Pi(x;\Omega_i)$ 是唯一的. 由于假设 $\bigcap_{i=1}^m \Omega_i = \varnothing$, 当 $i \in I(x)$ 时有 $\mathcal{M}(x) > 0$. 对任意的 $i \in I(x)$ 这自动保证 $x \notin \Omega_i$.

现基于上面建立的凸分析和最优化的一般结果给出问题 (4.55) 的必要和充分最优性条件.

定理 4.65 在所给的假设下元素 $\bar{x} \in \mathbb{R}^n$ 是问题 (4.55) 的一个最优解当且仅当有包含关系

$$\bar{x} \in \operatorname{co}\{\Pi(\bar{x};\Omega_i) \mid i \in I(\bar{x})\} \tag{4.57}$$

成立. 特别地, 对 $\Omega_i = \{a_i\}, i = 1, \cdots, m$ 的情形, 问题 (4.55) 有唯一解, 并且 $\bar{x}$ 是由 a_i 生成的该问题的唯一解当且仅当

$$\bar{x} \in \operatorname{co}\{a_i \mid i \in I(\bar{x})\}. \tag{4.58}$$

证明 由命题 2.35 和命题 2.54 知条件

$$0 \in \operatorname{co} \bigcup_{i \in I(\bar{x})} \partial d(\bar{x};\Omega_i)$$

对问题 (4.55) 中 $\bar{x}$ 的最优性来说是必要和充分的. 现通过式 (4.56) 中所计算的单点 $a_i(\bar{x})$ 利用定理 2.39 中距离函数的次微分给出极小值刻画

$$0 \in \operatorname{co}\{a_i(\bar{x}) \mid i \in I(\bar{x})\}. \tag{4.59}$$

易见式 (4.59) 成立当且仅当存在 $\lambda_i \geqslant 0, i \in I(\bar{x})$ 使得 $\sum\limits_{i \in I(\bar{x})} \lambda_i = 1$ 且

$$0 = \sum_{i \in I(\bar{x})} \lambda_i \frac{\bar{x} - \bar{\omega}_i}{\mathcal{M}(\bar{x})}, \quad \text{其中} \ \ \bar{\omega}_i := \Pi(\bar{x};\Omega_i).$$

以上方程显然等价于

$$0 \in \sum_{i \in I(\bar{x})} \lambda_i(\bar{x} - \bar{\omega}_i) \quad \text{或} \quad \bar{x} = \sum_{i \in I(\bar{x})} \lambda_i \bar{\omega}_i,$$

这等价于包含关系 (4.57). 最后, 对 $\Omega_i = \{a_i\}$, $i = 1, \cdots, m$, 包含关系 (4.57) 显然简化为条件 (4.58). 此时根据推论 4.64 问题有唯一最优解. □

为描述下面的结果, 称球 $\mathbb{B}(\bar{x};r)$ 与集合 Ω_i 相切, 如果交集 $\Omega_i \cap \mathbb{B}(\bar{x};r)$ 是单点集.

推论 4.66 在上面的假设下考虑问题 (4.55), 则任何最小相交球与目标集合 $\Omega_i, i = 1, \cdots, m$ 中至少有两个相触.

证明 首先证明如果 $\bar{x}$ 是问题 (4.55) 的解, 则至少存在 $I(\bar{x})$ 中的两个指标. 假设相反, 即 $I(\bar{x})=\{i_0\}$, 则由式 (4.57) 得

$$\bar{x}\in\Pi(\bar{x};\Omega_{i_0}) \text{ 且 } d(\bar{x};\Omega_{i_0})=\mathcal{M}(\bar{x})>0,$$

矛盾. 下证当 $i\in I(\bar{x})$ 时 $\mathbb{B}(\bar{x};r)$, 其中 $r:=\mathcal{M}(\bar{x})$, 与 Ω_i 相触. 事实上, 如果相反存在 $u,v\in\Omega_i$ 使得

$$u,v\in\mathbb{B}(\bar{x};r)\cap\Omega_i, \text{ 且 } u\neq v,$$

则根据 $d(\bar{x};\Omega_i)=r$ 得

$$\|u-\bar{x}\|\leqslant r=d(\bar{x};\Omega_i), \quad \|v-\bar{x}\|\leqslant r=d(\bar{x};\Omega_i),$$

因此 $u,v\in\Pi(\bar{x};\Omega_i)$. 这与投影 $\Pi(\bar{x};\Omega_i)$ 是单点集的事实矛盾, 从而完成了推论的证明. □

现在通过解下面两个例子来说明所得结果.

例 4.67 设 $\bar{x}$ 是由 $a_i\in\mathbb{R}^2, i=1,2,3$ 生成的最小封闭球问题的解. 由推论 4.66 知存在两个或三个活跃指标. 如果 $I(\bar{x})$ 由两个指标组成, 比如, $I(\bar{x})=\{2,3\}$, 则根据定理 4.65 可知, $\bar{x}\in\operatorname{co}\{a_2,a_3\}, \|\bar{x}-a_2\|=\|\bar{x}-a_3\|>\|\bar{x}-a_1\|$. 此时有 $\bar{x}=\dfrac{a_2+a_3}{2}, \langle a_2-a_1,a_3-a_1\rangle<0$. 反之, 如果 $\langle a_2-a_1,a_3-a_1\rangle<0$, 则以 a_1 为顶点由 a_1,a_2 与 a_3 构成的三角形的在顶点 a_1 处的角是钝角; 这也包含所有三个点 a_i 在一条直线上的情形. 易见 $I\left(\dfrac{a_2+a_3}{2}\right)=\{2,3\}$, 且点 $\bar{x}=\dfrac{a_2+a_3}{2}$ 满足定理 4.65 的假设从而是问题的解. 注意到此时问题 (4.55) 的解是钝角顶点对边的中点.

如果由 a_1,a_2 与 a_3 构成的三角形没有一个角是钝角, 则活跃指标集 $I(\bar{x})$ 由三个元素组成. 此时 $\bar{x}$ 是唯一满足

$$\bar{x}\in\operatorname{co}\{a_1,a_2,a_3\} \text{ 且 } \|\bar{x}-a_1\|=\|\bar{x}-a_2\|=\|\bar{x}-a_3\|$$

的点, 即 $\bar{x}$ 是三角形外接圆的圆心.

例 4.68 设 $\Omega_i=\mathbb{B}(c_i,r_i), r_i>0, i=1,2,3$ 是 $\mathbb{R}^2$ 中互不相交的圆盘, 设 $\bar{x}$ 是相应问题 (4.55) 的唯一解. 首先考虑连接其中两个圆心的线段与第三个圆盘相交的情形. 例如, 连接 c_2 与 c_3 的线段与 Ω_1 相交. 记 $u_2:=\overline{c_2c_3}\cap\operatorname{bd}\Omega_2, u_3:=\overline{c_2c_3}\cap\operatorname{bd}\Omega_3$, 设 $\bar{x}$ 是 $\overline{u_2u_3}$ 的中点, 则 $I(\bar{x})=\{2,3\}$, 并且利用定理 4.65 得 $\bar{x}$ 是问题的解 (图 4.9).

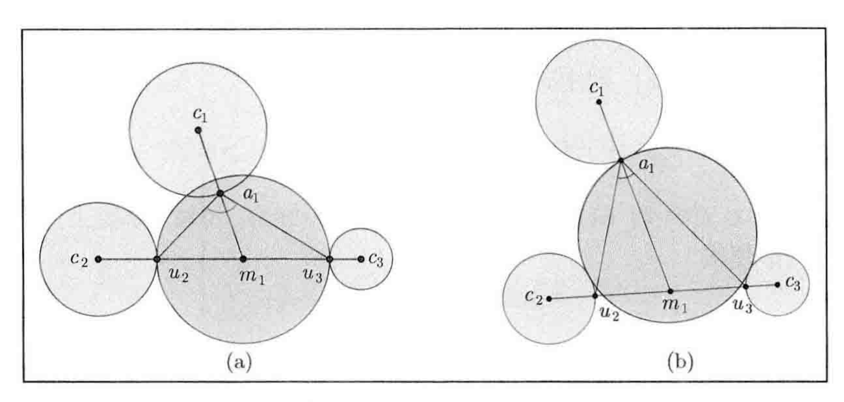


图 4.9　关于三个圆盘的最小相交圆问题

余下考虑任何连接两个圆盘圆心的线段不与剩余圆盘相交的情形. 此时记

$$\begin{aligned}
u_1 &:= \overline{c_1c_2} \cap \operatorname{bd}\Omega_1, \quad v_1 := \overline{c_1c_3} \cap \operatorname{bd}\Omega_1,\\
u_2 &:= \overline{c_2c_3} \cap \operatorname{bd}\Omega_2, \quad v_2 := \overline{c_2c_1} \cap \operatorname{bd}\Omega_2,\\
u_3 &:= \overline{c_3c_1} \cap \operatorname{bd}\Omega_3, \quad v_3 := \overline{c_3c_2} \cap \operatorname{bd}\Omega_3,
\end{aligned}$$

$$a_1 := \overline{c_1m_1} \cap \operatorname{bd}\Omega_1, \quad a_2 := \overline{c_2m_2} \cap \operatorname{bd}\Omega_2, \quad a_3 := \overline{c_3m_3} \cap \operatorname{bd}\Omega_3,$$

这里 m_1 是 $\overline{u_2u_3}$ 的中点, m_2 是 $\overline{u_3u_1}$ 的中点, m_3 是 $\overline{u_1u_2}$ 的中点. 假设三个角 $\widehat{u_2a_1v_3}$, $\widehat{u_3a_2v_1}$, $\widehat{u_1a_3v_2}$ 中有一个角大于或等于 90°. 例如, $\widehat{u_2a_1v_3}$ 大于 90°, 则 $I(m_1) = \{2,3\}$, 从而根据定理 4.65 可知, $\bar{x} = m_1$ 是问题的唯一解. 如果假设前面提到的所有角小于或等于 90°, 则 $I(\bar{x}) = 3$, 且所寻找的最小圆盘是与其他三个圆盘相触的唯一圆盘. 如何构造该圆盘是著名的 Apollonius 问题.

本节最后给出并举例说明用次梯度算法来数值求解广义 Sylvester 问题 (4.52).

定理 4.69　在问题 (4.52) 的情形下假设问题的解集 S 是非空的. 取定 $x_1 \in \Omega_0$, 并定义迭代序列为

$$x_{k+1} := \Pi(x_k - \alpha_k v_k; \Omega_0), \quad k \in \mathbb{N},$$

这里, $\{\alpha_k\}$ 是正数序列, 而

$$v_k \in \begin{cases} \{0\}, & x_k \in \Omega_i, \\ [-\partial\rho_F(\omega_k - x_k)] \cap N(\omega_k; \Omega_i), & x_k \notin \Omega_i, \end{cases}$$

而且对某 $i \in I(x_k)$, $\omega_k \in \Pi_F(x_k; \Omega_i)$. 定义值序列

$$V_k := \min\{\mathcal{T}(x_j) | j = 1, \cdots, k\}.$$

(i) 如果序列 $\{\alpha_k\}$ 满足条件 (4.13), 则 $\{V_k\}$ 收敛于问题的最优值 $\overline{V}$. 进一步, 有估计

$$0 \leqslant V_k - \overline{V} \leqslant \frac{d(x_1;S)^2 + \ell^2 \sum_{k=1}^{\infty} \alpha_k^2}{2\sum_{i=1}^{k} \alpha_k},$$

其中 $\ell \geqslant 0$ 是费用函数 $\mathcal{T}(\cdot)$ 在 $\mathbb{R}^n$ 上的 Lipschitz 常数.

(ii) 如果序列 $\{\alpha_k\}$ 满足条件 (4.15), 则 $\{V_k\}$ 收敛于最优值 $\overline{V}$, 且式 (4.49) 中 $\{x_k\}$ 收敛于问题的最优解 $\bar{x} \in S$.

证明 将 4.3 节的结果应用于问题 (4.52) 即可得如下结论. □

下面的例子说明定理 4.69 中的次梯度算法在特殊情形下的应用.

例 4.70 考虑 $\mathbb{R}^2$ 中的问题 (4.55), 其中 m 个目标集由正方形

$$\Omega_i = S(\omega_i;r_i) := [\omega_{1i} - r_i,\ \omega_{1i} + r_i] \times [\omega_{2i} - r_i,\ \omega_{2i} + r_i], \quad i = 1, \cdots, m$$

给出. 记第 i 个正方形的顶点为 $v_{1i} := (\omega_{1i} + r_i, \omega_{2i} + r_i), v_{2i} := (\omega_{1i} - r_i, \omega_{2i} + r_i)$, $v_{3i} := (\omega_{1i} - r_i, \omega_{2i} - r_i), v_{4i} := (\omega_{1i} + r_i, \omega_{2i} - r_i)$, 并设 $x_k := (x_{1k}, x_{2k})$. 固定指标 $i \in I(x_k)$, 于是根据定理 4.69 可选取元素 v_k 为 (图 4.10)

$$v_k = \begin{cases} \dfrac{x_k - v_{1i}}{\|x_k - v_{1i}\|}, & x_{1k} - \omega_{1i} > r_i \text{ 且 } x_{2k} - \omega_{2i} > r_i, \\ \dfrac{x_k - v_{2i}}{\|x_k - v_{2i}\|}, & x_{1k} - \omega_{1i} < -r_i \text{ 且 } x_{2k} - \omega_{2i} > r_i, \\ \dfrac{x_k - v_{3i}}{\|x_k - v_{3i}\|}, & x_{1k} - \omega_{1i} < -r_i \text{ 且 } x_{2k} - \omega_{2i} < -r_i, \\ \dfrac{x_k - v_{4i}}{\|x_k - v_{4i}\|}, & x_{1k} - \omega_{1i} > r_i \text{ 且 } x_{2k} - \omega_{2i} < -r_i, \\ (0,1), & |x_{1k} - \omega_{1i}| \leqslant r_i \text{ 且 } x_{2k} - \omega_{2i} > r_i, \\ (0,-1), & |x_{1k} - \omega_{1i}| \leqslant r_i \ \ x_{2k} - \omega_{2i} < -r_i, \\ (1,0), & x_{1k} - \omega_{1i} > r_i \text{ 且 } |x_{2k} - \omega_{2i}| \leqslant r_i, \\ (-1,0), & x_{1k} - \omega_{1i} < -r_i \text{ 且 } |x_{2k} - \omega_{2i}| \leqslant r_i. \end{cases}$$

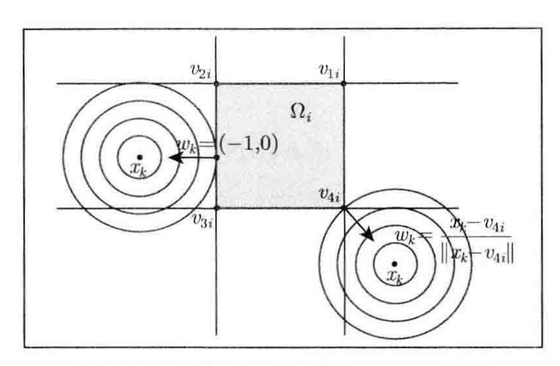


图 4.10　到一个正方形的 Euclid 投影

现考虑在极小时间函数中非 Euclid 范数的情形.

例 4.71　考虑问题 (4.52), 其中 $F := \{(x_1, x_2) \in \mathbb{R}^2 \mid |x_1| + |x_2| \leqslant 1\}$, 并具有非空闭凸目标集合 Ω. 中点在 $\bar{x} = (\bar{x}_1, \bar{x}_2)$ 半径为 $t > 0$ 的广义球 $G_F(\bar{x}; t)$ 的形状是下面的菱形 (图 4.11):

$$G_F(\bar{x}; t) = \{(x_1, x_2) \in \mathbb{R}^2 \mid |x_1 - \bar{x}_1| + |x_2 - \bar{x}_2| \leqslant t\}.$$

从 $\bar{x}$ 到 Ω 的距离和相应的投影由

$$\mathcal{T}_F(\bar{x}; \Omega) = \min\{t \geqslant 0 \mid G_F(\bar{x}; t) \cap \Omega \neq \varnothing\},$$

$$\Pi_F(\bar{x}; \Omega) = G_F(\bar{x}; t) \cap \Omega, \quad \text{其中} \quad t = \mathcal{T}_F(\bar{x}; \Omega) \tag{4.60}$$

给出. 设 Ω_i 是例 4.70 中给出的正方形 $S(\omega_i; r_i), i = 1, \cdots, m$, 则问题 (4.55) 可阐述如下: 找到 $\mathbb{R}^2$ 中的菱形集合与所有给定的 m 个正方形相交. 利用与例 4.70 中相同的记号, 可看到定理 4.69 中的元素 v_k 可选取为

$$v_k = \begin{cases} (1,1), & x_{1k} - \omega_{1i} > r_i \text{ 且 } x_{2k} - \omega_{2i} > r_i, \\ (-1,1), & x_{1k} - \omega_{1i} < -r_i \text{ 且 } x_{2k} - \omega_{2i} > r_i, \\ (-1,-1), & x_{1k} - \omega_{1i} < -r_i \text{ 且 } x_{2k} - \omega_{2i} < -r_i, \\ (1,-1), & x_{1k} - \omega_{1i} > r_i \text{ 且 } x_{2k} - \omega_{2i} < -r_i, \\ (0,1), & |x_{1k} - \omega_{1i}| \leqslant r_i \text{ 且 } x_{2k} - \omega_{2i} > r_i, \\ (0,-1), & |x_{1k} - \omega_{1i}| \leqslant r_i \text{ 且 } x_{2k} - \omega_{2i} < -r_i, \\ (1,0), & x_{1k} - \omega_{1i} > r_i \text{ 且 } |x_{2k} - \omega_{2i}| \leqslant r_i, \\ (-1,0), & x_{1k} - \omega_{1i} < -r_i \text{ 且 } |x_{2k} - \omega_{2i}| \leqslant r_i. \end{cases}$$

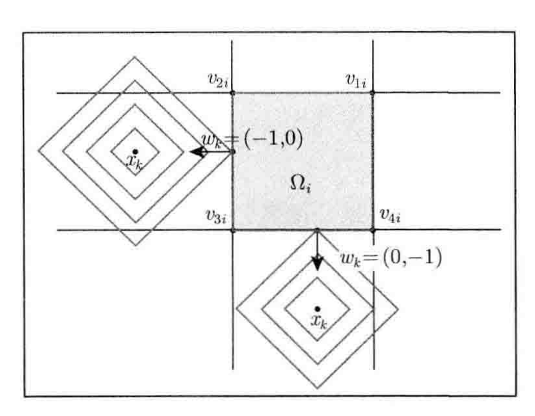


图 4.11　菱形动态的极小时间函数

4.7 练　习

练习 4.1　给定下半连续函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$, 证明它与任意常数 $\alpha>0$ 的乘积 $(\alpha f)(x)=\alpha f(x)$ 也是下半连续的.

练习 4.2　给定函数 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$, 证明它的下半连续性等价于对任意的 $\lambda\in\mathbb{R}$, 集合

$$\mathcal{U}_\lambda := \{x\in\mathbb{R}^n \mid f(x)>\lambda\}$$

的开性.

练习 4.3　给出两个下半连续实值函数的乘积不是下半连续的例子.

练习 4.4　设 $f:\mathbb{R}^n\to\mathbb{R}$, 证明下列的断言:

(i) 在定义 2.89 的意义下 f 在 $\bar{x}$ 处是上半连续 (u.s.c.) 的当且仅当 $-f$ 在该点是下半连续的.

(ii) f 在 $\bar{x}$ 处是 u.s.c. 的当且仅当 $\limsup_{k\to\infty} f(x_k)\leqslant f(\bar{x})$, 对任意的 $x_k\to\bar{x}$.

(iii) f 在 $\bar{x}$ 处是 u.s.c. 的当且仅当 $\limsup_{x\to\bar{x}} f(x)\leqslant f(\bar{x})$.

(iv)f 在 $\mathbb{R}^n$ 上是 u.s.c. 的当且仅当对任意实数 λ, 上水平集 $\{x\in\mathbb{R}^n \mid f(x)\geqslant\lambda\}$ 是闭的.

(v) f 在 $\mathbb{R}^n$ 上是 u.s.c. 的当且仅当它的图像下的集合 (或简称下图) $\operatorname{hypo} f := \{(x,\lambda)\in\mathbb{R}^n\times\mathbb{R} \mid \lambda\leqslant f(x)\}$ 是闭的.

练习 4.5　证明函数 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $\bar{x}$ 处是连续的当且仅当它在该点既是下半连续的, 又是上半连续的.

练习 4.6　证明定理 4.11 中的等价性 (iii)$\Leftrightarrow$(iv).

练习 4.7 考虑实值函数

$$g(x) := \sum_{i=1}^{m} \|x - a_i\|^2, \quad x \in \mathbb{R}^n,$$

这里 $a_i \in \mathbb{R}^n, i = 1, \cdots, m$. 证明 g 是凸函数且在几何平均点 $\bar{x} = \dfrac{a_1 + a_2 + \cdots + a_m}{m}$ 处有绝对极小值.

练习 4.8 考虑问题

$$\min \|x\|^2, \quad \text{s.t.} \quad Ax = b,$$

这里 A 是 $p \times n$ 矩阵且 A 的秩 $\text{rank}(A) = p$, $b \in \mathbb{R}^p$. 找到这个问题的唯一最优解.

练习 4.9 不利用定理 4.14, 给出推论 4.15 的一个直接证明.

练习 4.10 给定 $a_i \in \mathbb{R}^n, i = 1, \cdots, m$, 写一个 Matlab 程序来实施次梯度算法, 其中欲最小化的函数 f 给出如下:

(i) $f(x) := \sum\limits_{i=1}^{m} \|x - a_i\|, \ x \in \mathbb{R}^n$.

(ii) $f(x) := \sum\limits_{i=1}^{m} \|x - a_i\|_1, \ x \in \mathbb{R}^n$.

(iii) $f(x) := \sum\limits_{i=1}^{m} \|x - a_i\|_\infty, x \in \mathbb{R}^n$.

练习 4.11 写一个 Matlab 程序实施投影次梯度算法来解问题:

$$\min \|x\|_1, \quad \text{s.t.} \quad Ax = b,$$

这里 A 是 $p \times n$ 矩阵, 且 A 的秩 $\text{rank}(A) = p, b \in \mathbb{R}^p$.

练习 4.12 在保持 4.3 节中的假设下考虑最优化问题 (4.9) 和迭代过程 (4.10). 假设最优值 $\overline{V}$ 已知, 而且序列 $\{\alpha_k\}$ 由

$$\alpha_k: = \begin{cases} 0, & 0 \in \partial f(x_k), \\ \dfrac{f(x_k) - \overline{V}}{\|v_k\|^2}, & \text{其他} \end{cases} \tag{4.61}$$

给出, 这里 $v_k \in \partial f(x_k)$. 证明:

(i) $f(x_k) \to \overline{V}(k \to \infty)$.

(ii) $0 \leqslant V_k - \overline{V} \leqslant \dfrac{\ell d(x_1; S)}{\sqrt{k}}, k \in \mathbb{N}$.

练习 4.13 设 $\Omega_i, i = 1, \cdots, m$ 是 $\mathbb{R}^n$ 的非空闭凸子集, 它们具有非空的交. 对函数

$$f(x) := \max\{d(x; \Omega_i) \mid i = 1, \cdots, m\}$$

应用迭代过程 (4.10) 来得到寻找 $\bar{x} \in \bigcap_{i=1}^{m} \Omega_i$ 的一个算法, 其中序列 $\{\alpha_k\}$ 由式 (4.61) 给出.

练习 4.14 由定理 4.14 中的条件 (ii) 得到推论 4.20 的结果.

练习 4.15 描述并证明定理 4.27 关于投影次梯度算法 (4.20) 的对应版本并用来解约束最优化问题 (4.19).

练习 4.16 证明引理 4.31 中反向的蕴涵关系.

练习 4.17 写一个 Matlab 程序来实施 Weiszfeld 算法.

练习 4.18 设 $a_i \in \mathbb{R}, i = 1, \cdots, m$, 求解问题

$$\min \sum_{i=1}^{m} |x - a_i|, \quad x \in \mathbb{R}.$$

练习 4.19 给出定理 4.51 的一个简化证明.

练习 4.20 证明定理 4.52 中的断言 (ii).

练习 4.21 给出定理 4.56 的一个完整证明.

练习 4.22 设 $\Omega_i := [a_i, b_i] \subset \mathbb{R}, i = 1, \cdots, m$ 是 m 个互不相交的区间. 求解下面的问题:

$$\min f(x) := \sum_{i=1}^{m} d(x; \Omega_i), \quad x \in \mathbb{R}.$$

练习 4.23 写一个 Matlab 程序来实施求解下面问题的次梯度算法:

$$\min f(x), \quad x \in \mathbb{R}^n,$$

这里 $f(x) := \sum_{i=1}^{m} d(x; \Omega_i)$, 其中 $\Omega_i := \mathbb{B}(c_i; r_i), i = 1, \cdots, m$.

练习 4.24 完成例 4.58 中问题的解. 写一个 Matlab 程序来实施这个算法.

练习 4.25 在 F 是 $\mathbb{R}^n$ 的闭单位球, 目标集合是 $\mathbb{R}^n$ 的闭 Euclid 球的情形下给出定理 4.63 的一个简化证明.

练习 4.26 在 F 是 $\mathbb{R}^n$ 的闭单位球的情形下给出定理 4.63 的一个简化证明.

练习 4.27 写一个 Matlab 程序来实施求解问题

$$\min f(x), \quad x \in \mathbb{R}^n$$

的次梯度算法, 这里 $f(x) := \max\{d(x; \Omega_i) \mid i = 1, \cdots, m\}$, 其中 $\Omega_i := \mathbb{B}(c_i; r_i), i = 1, \cdots, m$. 注意到如果 $\bar{x}$ 是该问题的最优解且具最优值为 r, 则 $\mathbb{B}(\bar{x}; r)$ 是与 $\Omega_i, i = 1, \cdots, m$ 相交的最小球.

练习 4.28 给出定理 4.69 的完整证明.

练习 4.29 解出例 4.70 中问题的解. 写一个 Matlab 程序来实施这个算法.

练习 4.30 解出例 4.71 中问题的解. 写一个 Matlab 程序来实施这个算法.

部分练习答案和提示

本书的最后部分包含某些练习的提示. 对几个问题给出了相当详细的解答.

第 1 章练习

练习 1.1 假设 Ω_1 是有界的. 取定序列 $\{x_k\} \subset \Omega_1 - \Omega_2$ 收敛于 $\bar{x}$. 对任意的 $k \in \mathbb{N}, x_k$ 可表示为 $x_k = y_k - z_k$, 其中 $y_k \in \Omega_1, z_k \in \Omega_2$. 由于 Ω_1 是闭有界的, $\{y_k\}$ 有子序列 $\{y_{k_\ell}\}$ 收敛于 $\bar{y} \in \Omega_1$, 则 $\{z_{k_\ell}\}$ 收敛于 $\bar{y} - \bar{x} \in \Omega_2$. 于是 $\bar{x} \in \Omega_1 - \Omega_2$. 从而 Ω 是闭的. 设 $\Omega_1 := \{(x,y) \in \mathbb{R}^2 \mid y \geqslant 1/x, x > 0\}, \Omega_2 := \{(x,y) \in \mathbb{R}^2 \mid y \leqslant -1/x, x > 0\}$, 则 Ω_1 与 Ω_2 是非空闭凸集合, 而 $\Omega_1 - \Omega_2 = \{(x,y) \in \mathbb{R}^2 \mid y > 0\}$ 不是闭的.

练习 1.4 (i) 利用练习 1.3(ii).

(ii) 我们仅考虑当 $m=2$ 的情形, 一般的情形是类似的, 由锥的定义, 集合 Ω_1 和 Ω_2 都包含 0, 因此 $\Omega_1 \cup \Omega_2 \subset \Omega_1 + \Omega_2$. 于是 $\text{co}\{\Omega_1 \cup \Omega_2\} \subset \Omega_1 + \Omega_2$. 为证反向包含关系成立, 任取 $x = w_1 + w_2 \in \Omega_1 + \Omega_2$ 且 $w_1 \in \Omega_1, w_2 \in \Omega_2$, 于是 $x = [(2w_1 + 2w_2)/2] \in \text{co}\{\Omega_1 \cup \Omega_2\}$.

练习 1.7 计算每个函数的一阶或二阶导数. 再应用定理 1.45 或推论 1.46.

练习 1.8 (i) 利用 $f(x,y) := xy, (x,y) \in \mathbb{R}^2$.

(ii) 取 $f_1(x) := x, f_2(x) := -x$.

练习 1.9 设 $f_1(x) := x, f_2(x) := x^2, x \in \mathbb{R}$.

练习 1.12 利用命题 1.39. 注意到函数 $f(x) := x^q$ 在 $[0,\infty)$ 上是凸的.

练习 1.16 定义 $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ 为 $F(x) := K, \forall x \in \mathbb{R}^n$, 则 $\text{gph}\, F$ 是 $\mathbb{R}^n \times \mathbb{R}^p$ 的凸子集. 结论由命题 1.50 可得.

练习 1.18 (i)$\text{gph}\, F = C$.

(ii) 考虑函数

$$f(x):\ = \begin{cases} -\sqrt{1-x^2}, & |x| \leqslant 1, \\ \infty, & \text{其他}. \end{cases}$$

(iii) 应用命题 1.50.

练习 1.20 (i) 利用定义. (ii) 取 $f(x) := \sqrt{|x|}$.

练习 1.22 假设 $\lambda v + \sum_{i=1}^m \lambda_i v_i = 0$ 且 $\lambda + \sum_{i=1}^m \lambda_i = 0$. 如果 $\lambda \neq 0$, 则 $-\sum_{i=1}^m \frac{\lambda_i}{\lambda} = 1$, 因此 $v = -\sum_{i=1}^m \frac{\lambda_i}{\lambda} v_i \in \text{aff}\{v_1, \cdots, v_m\}$, 这是个矛盾. 所以 $\lambda = 0$, 从而 $\lambda_i = 0, i = 1, \cdots, m$, 因为元素 $v_1, \cdots, v_m$ 是仿射无关的.

练习 1.23 由于 $\Delta_m \subset \Omega \subset \operatorname{aff}\Omega$, 有 $\operatorname{aff}\Delta_m \subset \operatorname{aff}\Omega$. 为证反向的包含关系成立, 余下需证 $\Omega \subset \operatorname{aff}\Delta_m$. 用反证法, 假设存在 $v \in \Omega$ 使得 $v \notin \operatorname{aff}\Delta_m$. 由练习 1.22 得向量 $v, v_1, \cdots, v_m$ 是仿射无关的, 而且它们都属于 Ω, 这与条件 $\dim\Omega = m$ 矛盾. 第二个等式也可类似地证明.

练习 1.24 (i) 显然有 $\operatorname{aff}\Omega \subset \operatorname{aff}\overline{\Omega}$. 注意到集合 $\operatorname{aff}\Omega$ 是闭的且包含 Ω, 这是由于 $\operatorname{aff}\Omega = w_0 + L$, 这里 $w_0 \in \Omega$, L 是平行于 $\operatorname{aff}\Omega$ 的线性子空间, 它是个闭集. 因此 $\overline{\Omega} \subset \operatorname{aff}\Omega$, 从而结论成立.

(ii) 选取 $\delta > 0$ 满足 $\mathbb{B}(\bar{x};\delta) \cap \operatorname{aff}\Omega \subset \Omega$. 由于 $\bar{x} + t(\bar{x} - x) = (1+t)\bar{x} + (-t)x$ 是 $\operatorname{aff}\Omega = \operatorname{aff}\overline{\Omega}$ 中一些元素的仿射组合, 有 $\bar{x} + t(\bar{x} - x) \in \operatorname{aff}\Omega$. 因此当 t 很小时 $u := \bar{x} + t(\bar{x} - x) \in \Omega$.

(iii) 由 (i) 得 $\operatorname{ri}\Omega \subset \operatorname{ri}\overline{\Omega}$. 任取 $x \in \operatorname{ri}\overline{\Omega}$ 且 $\bar{x} \in \operatorname{ri}\Omega$. 由 (ii), 当 t 很小时有 $z := x + t(x - \bar{x}) \in \overline{\Omega}$, 因此 $x = z/(1+t) + t\bar{x}/(1+t) \in \operatorname{ri}\Omega$.

练习 1.26 如果 $\overline{\Omega}_1 = \overline{\Omega}_2$, 则 $\operatorname{ri}\overline{\Omega}_1 = \operatorname{ri}\overline{\Omega}_2$, 因此由练习 1.24(iii) 得 $\operatorname{ri}\Omega_1 = \operatorname{ri}\Omega_2$.

练习 1.27 (i) 取 $y \in B(\operatorname{ri}\Omega)$ 并找到 $x \in \operatorname{ri}\Omega$ 使得 $y = B(x)$, 则取向量 $\bar{y} \in \operatorname{ri}B(\Omega) \subset B(\Omega)$, 并取 $\bar{x} \in \Omega$ 且 $\bar{y} \in B(\bar{x})$. 如果 $x = \bar{x}$, 则 $y = \bar{y} \in \operatorname{ri}B(\Omega)$. 考虑 $x \neq \bar{x}$ 的情形. 沿用练习 1.24(iii) 的解法, 找到 $\tilde{x} \in \Omega$ 使得 $x \in (\tilde{x}, \bar{x})$, 并设 $\tilde{y} := B(\tilde{x}) \in B(\Omega)$. 因此 $y \in B(x) \in (B(\tilde{x}), B(\bar{x})) = (\tilde{y}, \bar{y})$, 从而 $y \in \operatorname{ri}B(\Omega)$. 为完成证明, 余下证明 $\overline{B(\Omega)} = \overline{B(\operatorname{ri}\Omega)}$, 然后利用练习 1.26 得包含关系

$$\operatorname{ri}B(\Omega) = \operatorname{ri}B(\operatorname{ri}\Omega) \subset B(\operatorname{ri}\Omega).$$

由于 $\overline{B(\operatorname{ri}\Omega)} \subset \overline{B(\Omega)}$, B 的连续性和命题 1.73 蕴涵着

$$B(\Omega) \subset B(\overline{\Omega}) = B(\overline{\operatorname{ri}\Omega}) \subset \overline{B(\operatorname{ri}\Omega)}.$$

因此有 $\overline{B(\Omega)} \subset \overline{B(\operatorname{ri}\Omega)}$, 于是完成了证明.

(ii) 考虑定义在 $\mathbb{R}^n \times \mathbb{R}^n$ 上的 $B(x,y) := x - y$.

练习 1.28 (i) 下证

$$\operatorname{aff}(\operatorname{epi} f) = \operatorname{aff}(\operatorname{dom} f) \times \mathbb{R}. \tag{4.62}$$

任取 $(x,\gamma) \in \operatorname{aff}(\operatorname{epi} f)$, 则存在 $\lambda_i \in \mathbb{R}, (x_i, \gamma_i) \in \operatorname{epi} f, i = 1, \cdots, m$ 且 $\sum_{i=1}^{m} \lambda_i = 1$ 满足

$$(x,\gamma) = \sum_{i=1}^{m} \lambda_i (x_i, \gamma_i).$$

由于 $x_i \in \operatorname{dom} f, i=1,\cdots,m$, 有 $x=\sum_{i=1}^{m}\lambda_i x_i \in \operatorname{aff}(\operatorname{dom} f)$, 从而

$$\operatorname{aff}(\operatorname{epi} f) \subset \operatorname{aff}(\operatorname{dom} f)\times\mathbb{R}.$$

为证反向的包含关系成立, 任取 $(x,\gamma)\in\operatorname{aff}(\operatorname{dom} f)\times\mathbb{R}$, 并找到 $x_i\in\operatorname{dom} f$ 和 $\lambda_i\in\mathbb{R}, i=1,\cdots,m$ 满足 $\sum_{i=1}^{m}\lambda_i=1$ 且 $x=\sum_{i=1}^{m}\lambda_i x_i$. 进一步定义 $\alpha_i:=f(x_i)\in\mathbb{R}$ 且 $\alpha:=\sum_{i=1}^{m}\lambda_i\alpha_i$ 满足 $(x_i,\alpha_i)\in\operatorname{epi} f$ 和 $(x_i,\alpha_i+1)\in\operatorname{epi} f$. 因此

$$\sum_{i=1}^{m}\lambda_i(x_i,\alpha_i)=(x,\alpha)\in\operatorname{aff}(\operatorname{epi} f),$$

$$\sum_{i=1}^{m}\lambda_i(x_i,\alpha_i+1)=(x,\alpha+1)\in\operatorname{aff}(\operatorname{epi} f).$$

令 $\lambda:=\alpha-\gamma+1$, 得

$$(x,\gamma)=\lambda(x,\alpha)+(1-\lambda)(x,\alpha+1)\in\operatorname{aff}(\operatorname{epi} f),$$

这就证明了式 (4.62) 中反向的包含关系成立.

(ii) 参见文献 [9] 的命题 1.1.9 的证明.

练习 1.30 (i) $\Pi(x;\Omega)=x+\dfrac{b-\langle a,x\rangle}{\|a\|^2}a$.

(ii) $\Pi(x;\Omega)=x+A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}(b-Ax)$.

(iii) $\Pi(x;\Omega)=\max\{x,0\}$(分量最大值).

第 2 章练习

练习 2.1 (i) 利用命题 2.1 和事实 $tx\in\Omega,\quad\forall t\geqslant 0$.

(ii) 利用 (i) 并注意到对任意的 $x\in\Omega$ 有 $-x\in\Omega$.

练习 2.2 注意到集合 $\operatorname{ri}\Omega_i, i=1,2$ 是非空凸的. 根据定理 2.5, 存在 $0\neq v\in\mathbb{R}^n$ 满足

$$\langle v,x\rangle\leqslant\langle v,y\rangle,\quad\forall x\in\operatorname{ri}\Omega_1, y\in\operatorname{ri}\Omega_2.$$

取 $x_0\in\operatorname{ri}\Omega_1, y_0\in\operatorname{ri}\Omega_2$. 由定理 1.72(ii) 得对任意的 $x\in\Omega_1, y\in\Omega_2, t\in(0,1)$, 有 $x+t(x_0-x)\in\operatorname{ri}\Omega_1, y+t(y_0-y)\in\operatorname{ri}\Omega_2$. 于是对这些元素利用上面的不等式, 并令 $t\to 0^+$.

练习 2.3 答案将分成四步.

第一步 如果 L 是 $\mathbb{R}^n$ 的子空间包含非空凸集 Ω 且 $\bar{x} \notin \overline{\Omega}, \bar{x} \in L$, 则存在 $0 \neq v \in L$ 满足

$$\sup\{\langle v, x\rangle \mid x \in \Omega\} < \langle v, \bar{x}\rangle. \tag{4.63}$$

证明 根据命题 2.1, 找到 $w \in \mathbb{R}^n$ 满足

$$\sup\{\langle w, x\rangle \mid x \in \Omega\} < \langle w, \bar{x}\rangle. \tag{4.64}$$

把 $\mathbb{R}^n$ 表示为 $L \oplus L^{\perp}$, 这里 $L^{\perp} := \{u \in \mathbb{R}^n \mid \langle u, x\rangle = 0, \forall x \in L\}$, 则 $w = u + v$, 其中 $u \in L^{\perp}, v \in L$. 把 w 代入式 (4.64) 并考虑到对任意的 $x \in \Omega, \langle u, x\rangle = 0$ 和 $\langle u, \bar{x}\rangle = 0$, 得式 (4.63) 成立. 注意到由式 (4.63) 得 $v \neq 0$.

第二步 设 $\Omega \subset \mathbb{R}^n$ 是非空凸集满足 $0 \notin \operatorname{ri}\Omega$, 则集合 Ω 和 $\{0\}$ 可正常分离.

证明 $0 \notin \overline{\Omega}$ 的情形, 可严格分离 Ω 与 $\{0\}$, 结论显然成立. 现假设 $0 \in \overline{\Omega} \backslash \operatorname{ri}\Omega$. 令 $L := \operatorname{aff}\Omega = \operatorname{span}\Omega$, 由于 $0 \in \operatorname{aff}\Omega$ 其中后面的式子成立, 由 L 是闭的且 $\overline{\Omega} \subset L$ 的事实可得. 重复引理 2.4 的证明, 找到序列 $\{x_k\} \subset L$ 满足 $x_k \to 0$ 且对任意的 k, $x_k \notin \overline{\Omega}$. 由第一步知存在 $\{v_k\} \subset L$ 满足 $v_k \neq 0$ 和

$$\sup\{\langle v_k, x\rangle \mid x \in \Omega\} < \langle v_k, x_k\rangle.$$

由于可以用 $\|v_k\|$ 同除两边, 不失一般性, 假设 $\|v_k\| = 1, v_k \to v \in L$ 且 $\|v\| = 1$, 则

$$\sup\{\langle v, x\rangle \mid x \in \Omega\} \leqslant 0.$$

为完成这个证明, 余下需证存在 $x \in \Omega$ 满足 $\langle v, x\rangle < 0$. 用反证法, 假设 $\langle v, x\rangle \geqslant 0$, $\forall x \in \Omega$, 则在 Ω 上 $\langle v, x\rangle = 0$. 由于 $v \in L = \operatorname{aff}\Omega$, 可把 v 表示为 $v = \sum_{i=1}^{m} \lambda_i w_i$, 其中 $\sum_{i=1}^{m} \lambda_i = 1, w_i \in \Omega, i = 1, \cdots, m$. 这蕴涵着 $\|v\|^2 = \langle v, v\rangle = \sum_{i=1}^{m} \lambda_i \langle v, w_i\rangle = 0$, 这是一个矛盾.

第三步 设 $\Omega \subset \mathbb{R}^n$ 是非空凸集. 如果集合 Ω 和 $\{0\}$ 可正常分离，则有 $0 \notin \operatorname{ri}\Omega$.

证明 根据定义, 找到 $0 \neq v \in \mathbb{R}^n$ 满足

$$\sup\{\langle v, x\rangle \mid x \in \Omega\} \leqslant 0,$$

于是取 $\bar{x} \in \Omega$ 满足 $\langle v, \bar{x}\rangle < 0$. 用反证法, 假设 $0 \in \operatorname{ri}\Omega$, 则由练习 1.24(ii) 知存在 $t > 0$, 使得 $0 + t(0 - \bar{x}) = -t\bar{x} \in \Omega$. 因此 $\langle v, -t\bar{x}\rangle \leqslant 0$, 从而 $\langle v, \bar{x}\rangle \geqslant 0$, 矛盾.

第四步 证明练习 2.3 的结论成立.

证明　定义 $\Omega := \Omega_1 - \Omega_2$. 根据练习 1.27(ii) 有 $\operatorname{ri}\Omega_1 \cap \operatorname{ri}\Omega_2 = \varnothing$ 当且仅当

$$0 \notin \operatorname{ri}(\Omega_1 - \Omega_2) = \operatorname{ri}\Omega_1 - \operatorname{ri}\Omega_2.$$

于是结论由第二步和第三步可得.

练习 2.4　注意到 $\Omega_1 \cap [\Omega_2 - (0, \lambda)] = \varnothing, \quad \forall \lambda > 0$.

练习 2.5　利用定理 2.5 并利用练习 2.2 中类似的证明, 可证存在 $0 \neq v \in \mathbb{R}^n$ 满足

$$\langle v, x\rangle \leqslant \langle v, y\rangle, \quad \forall\, x \in \Omega_1, y \in \Omega_2.$$

根据命题 2.13, 选取 $0 \neq v \in N(\bar{x}; \Omega_1) \cap [-N(\bar{x}; \Omega_2)]$, 并利用法锥的定义, 令 $v_1 := v, v_2 := -v, \alpha_1 := \langle v_1, \bar{x}\rangle, \alpha_2 := \langle v_2, \bar{x}\rangle$, 则结论成立.

练习 2.6　取 $\bar{x} \leqslant 0, v \in N(\bar{x}; \Omega)$, 则

$$\langle v, x - \bar{x}\rangle \leqslant 0, \quad \forall x \in \Omega.$$

对 $0 \in \Omega$ 和 $2\bar{x} \in \Omega$ 应用这个不等式, 得 $\langle v, \bar{x}\rangle = 0$. 由于 $\Omega = (-\infty, 0] \times \cdots \times (-\infty, 0]$, 由命题 2.11 得 $v \geqslant 0$. 反向的包含关系由定义直接可得.

练习 2.8　假设对任意 k 个凸集这个结论成立, 其中 $k \leqslant m$ 且 $m \geqslant 2$. 考虑集合 $\Omega_i, i = 1, \cdots, m+1$, 假设它们满足规范条件

$$\left[\sum_{i=1}^{m+1} v_i = 0,\ v_i \in N(\bar{x}; \Omega_i)\right] \Rightarrow [v_i = 0, i = 1, \cdots, m+1].$$

由于 $0 \in N(\bar{x}; \Omega_{m+1})$, 这个条件对前 m 个集合也成立. 因此

$$N\left(\bar{x}; \bigcap_{i=1}^{m} \Omega_i\right) = \sum_{i=1}^{m} N(\bar{x}; \Omega_i).$$

接着应用归纳法假设于 $\Theta_1 := \bigcap_{i=1}^{m} \Omega_i$ 和 $\Theta_2 := \Omega_{m+1}$.

练习 2.10　由于 $v \in \partial f(\bar{x})$, 有

$$v(x - \bar{x}) \leqslant f(x) - f(\bar{x}), \quad \forall x \in \mathbb{R}.$$

于是任取 $x_1 < \bar{x}$ 并应用上面的不等式.

练习 2.12　通过 $\{\bar{x}\}$ 的指示函数定义函数

$$g(x) := f(x) + \delta_{\{\bar{x}\}}(x) = \begin{cases} f(x), & x = \bar{x}, \\ \infty, & \text{其他}, \end{cases}$$

则 $\operatorname{epi} g = \{\bar{x}\} \times [f(\bar{x}), \infty)$, 因此 $N((\bar{x}, g(\bar{x})); \operatorname{epi} g) = \mathbb{R}^n \times (-\infty, 0]$. 由命题 2.31 得 $\partial g(\bar{x}) = \mathbb{R}^n$ 和 $\partial \delta_{\{\bar{x}\}} = \mathbb{R}^n$. 利用推论 2.45 中的次微分和法则得

$$\mathbb{R}^n = \partial g(\bar{x}) = \partial f(\bar{x}) + \mathbb{R}^n,$$

这保证 $\partial f(\bar{x}) \neq \varnothing$.

练习 2.13 像定理 2.15 的证明那样一步步进行. 由练习 1.28 得 $\operatorname{ri}\Theta_1 = \operatorname{ri}\Omega_1 \times (0, \infty)$, 而且

$$\operatorname{ri}\Theta_2 = \{(x, \lambda) \mid x \in \operatorname{ri}\Omega_1, \lambda < \langle v, x - \bar{x} \rangle\}.$$

利用反证法, 易验证 $\operatorname{ri}\Theta_1 \cap \operatorname{ri}\Theta_2 = \varnothing$. 利用练习 2.3 中的分离结果, 可找到 $0 \neq (w, \gamma) \in \mathbb{R}^n \times \mathbb{R}$ 满足

$$\langle w, x \rangle + \lambda_1 \gamma \leqslant \langle w, y \rangle + \lambda_2 \gamma, \quad \forall (x, \lambda_1) \in \Theta_1, (y, \lambda_2) \in \Theta_2.$$

而且, 存在 $(\tilde{x}, \tilde{\lambda}_1) \in \Theta_1, (\tilde{y}, \tilde{\lambda}_2) \in \Theta_2$ 满足

$$\langle w, \tilde{x} \rangle + \tilde{\lambda}_1 \gamma < \langle w, \tilde{y} \rangle + \tilde{\lambda}_2 \gamma.$$

与定理 2.15 的证明相同, 余下证明 $\gamma < 0$. 用反证法, 假设 $\gamma = 0$, 则得

$$\langle w, x \rangle \leqslant \langle w, y \rangle, \quad \forall x \in \Omega_1, y \in \Omega_2,$$

$$\langle w, \tilde{x} \rangle \leqslant \langle w, \tilde{y} \rangle, \quad \tilde{x} \in \Omega_1, \tilde{y} \in \Omega_2.$$

于是集合 Ω_1 和 Ω_2 是正常分离的, 因此由练习 2.3 得 $\operatorname{ri}\Omega_1 \cap \operatorname{ri}\Omega_2 = \varnothing$. 这是一个矛盾, 从而证明了 $\gamma < 0$ 成立. 为完成证明, 只需重复定理 2.15 中情形 2 的证明.

练习 2.14 利用练习 2.13, 并仿照定理 2.44 的证明.

练习 2.16 利用与定理 2.51 的证明相同的过程和奇异次微分的结构, 得

$$\partial^\infty (f \circ B)(\bar{x}) = A^*(\partial^\infty f(\bar{y})) = \{A^* v \mid v \in \partial^\infty f(\bar{y})\}. \tag{4.65}$$

作为另一办法, 也可注意到等式 $\operatorname{dom}(f \circ B) = B^{-1}(\operatorname{dom} f)$, 然后利用命题 2.23 和推论 2.53 来完成证明.

练习 2.16 利用定理 2.61 中的次微分公式, 接着利用命题 1.54 中复合的表达式.

练习 2.19 对任意的 $v \in \mathbb{R}^n$, 有 $\sigma_{\Omega_1}(v) = \|v\|_\infty$ 和 $\sigma_{\Omega_2}(v) = \|v\|_1$.

练习 2.21 首先利用表达式

$$(\delta_\Omega)^*(v) = \sup\{\langle v, x \rangle - \delta(x; \Omega) \mid x \in \mathbb{R}^n\} = \sup\{\langle v, x \rangle \mid x \in \Omega\} = \sigma_\Omega(v), \quad v \in \mathbb{R}^n.$$

接着再次利用 Fenchel 共轭得

$$(\sigma_\Omega)^*(x) = \sup\{\langle x, v\rangle - \sigma_\Omega(v) \mid v \in \mathbb{R}^n\} = \delta_{\overline{\text{co}}\Omega}(x).$$

练习 2.23 由于 $f:\mathbb{R}^n \to \mathbb{R}$ 是凸正齐次的, 有 $\partial f(0) \neq \varnothing$ 和 $f(0) = 0$. 这蕴涵着当 $v \in \Omega = \partial f(0)$ 时有 $\langle v, x\rangle \leqslant f(x), \forall x \in \mathbb{R}^n$, 因此当 $x = 0$ 时等式成立. 从而有

$$f^*(v) = \sup\{\langle v, x\rangle - f(x) \mid x \in \mathbb{R}^n\} = 0.$$

如果 $v \notin \Omega = \partial f(0)$, 则存在 $\bar{x} \in \mathbb{R}^n$ 满足 $\langle v, \bar{x}\rangle > f(\bar{x})$. 根据齐次性得

$$f^*(v) = \sup\{\langle v, x\rangle - f(x) \mid x \in \mathbb{R}^n\} \geqslant \sup\{\langle v, t\bar{x}\rangle - f(t\bar{x}) \mid t > 0\} = \infty.$$

练习 2.26 $f'(-1; d) = -\infty, f'(1; d) = \infty$.

练习 2.27 对固定的 $t_1 < t_2 < 0$, 有

$$\bar{x} + t_2 d = \frac{t_2}{t_1}(\bar{x} + t_1 d) + \left(1 - \frac{t_2}{t_1}\right)\bar{x}.$$

练习 2.28 (i) 对 $d \in \mathbb{R}^n$, 考虑引理 2.80 中定义的函数 φ, 然后证明不等式 $\lambda < 0 < t$ 蕴涵着 $\varphi(\lambda) \leqslant \varphi(t)$.

(ii) 由 (i) 得 $\varphi(t) \leqslant \varphi(1), \forall t \in (0, 1)$. 同理, $\varphi(-1) \leqslant \varphi(t), \forall t \in (-1, 0)$. 由此易得结论成立.

练习 2.31 取定 $\bar{x} \in \mathbb{R}^n$, 定义 $\mathcal{P}(\bar{x}; \Omega) := \{w \in \Omega \mid \|\bar{x} - w\| = \mu_\Omega(\bar{x})\}$, 考虑下面的函数

$$f_w(x) := \|x - w\|, \quad w \in \Omega, \quad x \in \mathbb{R}^n,$$

则由定理 2.93 得

$$\partial\mu_\Omega(\bar{x}) = \text{co}\,\{\partial f_w(\bar{x}) \mid w \in \mathcal{P}(\bar{x}; \Omega)\}.$$

当 Ω 不是单点集时, 有 $\partial f_w(\bar{x}) = \dfrac{\bar{x} - w}{\mu_\Omega(x)}$, 因此

$$\partial\mu_\Omega(\bar{x}) = \frac{\bar{x} - \text{co}\,\mathcal{P}(\bar{x}; \Omega)}{\mu_\Omega(\bar{x})},$$

Ω 是单点集的情形是平凡的.

练习 2.33 (i) 任取 $v \in \mathbb{R}^n$ 且 $\|v\| \leqslant 1$, 任取 $w \in \Omega$, 则

$$\langle x, v\rangle - \sigma_\Omega(v) \leqslant \langle x, v\rangle - \langle w, v\rangle = \langle x - w, v\rangle \leqslant \|x - w\| \cdot \|v\| \leqslant \|x - w\|.$$

从而有

$$\sup_{\|v\| \leqslant 1}\{\langle x, v\rangle - \sigma_\Omega(v)\} \leqslant d(x; \Omega).$$

为证反向的不等式成立, 首先注意到如果 $x \in \Omega$, 它平凡成立. 当 $x \notin \Omega$ 时, 令 $w := \Pi(x;\Omega), u := x - w$; 于是 $\|u\| = d(x;\Omega)$. 沿用命题 2.1 的证明, 得估计

$$\langle u, z\rangle \leqslant \langle u, x\rangle - \|u\|^2, \quad \forall\, z \in \Omega.$$

用 $\|u\|$ 同除上式两边, 并令 $v := u/\|u\|$, 则 $\sigma_\Omega(v) \leqslant \langle v, x\rangle - d(x;\Omega)$, 因此

$$d(x;\Omega) \leqslant \sup_{\|v\|\leqslant 1} \{\langle x, v\rangle - \sigma_\Omega(v)\}.$$

第 3 章练习

练习 3.1 假设 f 在 $\bar{x}$ 处是 Gâteaux 可微的, 且 $f'_G(\bar{x}) = v$, 则

$$\langle f'_G(\bar{x}), d\rangle = \lim_{t\to 0}\frac{f(\bar{x}+td) - f(\bar{x})}{t} = \lim_{t\to 0^+}\frac{f(\bar{x}+td) - f(\bar{x})}{t}, \quad \forall\, d \in \mathbb{R}^n.$$

由方向导数的定义上式蕴涵着 $f'(\bar{x};d) = \langle v, d\rangle, \forall\, d \in \mathbb{R}^n$. 反之, 假设 $f'(\bar{x};d) = \langle v, d\rangle, \forall\, d \in \mathbb{R}^n$, 则

$$f'(\bar{x};d) = \langle v, d\rangle = \lim_{t\to 0^+}\frac{f(\bar{x}+td) - f(\bar{x})}{t}, \quad \forall\, d \in \mathbb{R}^n.$$

而且,

$$f'_-(\bar{x};d) = -f'(\bar{x};-d) = -\langle v, -d\rangle = \langle v, d\rangle, \quad \forall\, d \in \mathbb{R}^n.$$

因此

$$f'_-(\bar{x};d) = \langle v, d\rangle = \lim_{t\to 0^-}\frac{f(\bar{x}+td) - f(\bar{x})}{t}, \quad \forall\, d \in \mathbb{R}^n.$$

这蕴涵着 f 在 $\bar{x}$ 处是 Gâteaux 可微的且 $f'_G(\bar{x}) = v$.

练习 3.2 (i) 由于函数 $\varphi(t) := t^2$ 在 $[0,\infty)$ 上是非减的, 沿用推论 2.62 的证明, 可证明如果 $g : \mathbb{R}^n \to [0,\infty)$ 是凸函数, 则 $\partial(g^2)(\bar{x}) = 2g(\bar{x})\partial g(\bar{x}), \forall\, \bar{x} \in \mathbb{R}^n$, 这里 $g^2(x) := [g(x)]^2$. 由定理 2.39 得

$$\partial f(\bar{x}) = 2d(\bar{x};\Omega)\partial d(\bar{x};\Omega) = \begin{cases} \{0\}, & \bar{x} \in \Omega, \\ \{2[\bar{x} - \Pi(\bar{x};\Omega)]\}, & \text{其他}. \end{cases}$$

因此在任何情形下 $\partial f(\bar{x}) = \{2[\bar{x} - \Pi(\bar{x};\Omega)]\}$ 是单点集, 根据定理 3.3, 这蕴涵着 f 在 $\bar{x}$ 处的可微性, 而且 $\nabla f(\bar{x}) = 2[\bar{x} - \Pi(\bar{x};\Omega)]$.

(ii) 在集合外的情形应用定理 2.39 和定理 3.3.

练习 3.3 对任意的 $u \in F$, 函数 $f_u(x) := \langle Ax, u\rangle - \dfrac{1}{2}\|u\|^2$ 是凸的, 从而根据命题 1.43 可知, f 是凸函数. 由等式 $\langle Ax, u\rangle = \langle x, A^*u\rangle$ 得

$$\nabla f_u(x) = A^*u.$$

取定 $x \in \mathbb{R}^n$, 并注意到函数 $g_x(u) := \langle Ax, u\rangle - \frac{1}{2}\|u\|^2$ 在 F 上有唯一极大值点, 记为 $u(x)$. 根据定理 2.93, 有 $\partial f(x) = \{A^*u(x)\}$, 因此, 根据定理 3.3 可知, f 在 x 处是可微的. 更多一般的结果及其在最优化中的应用建议读者参阅文献 [22].

练习 3.5 考虑集合 $Y := \{d \in \mathbb{R}^n \mid f'(\bar{x}; d) = f'_-(\bar{x}; d)\}$, 下证 Y 是包含标准正交基 $\{e_i \mid i = 1, \cdots, n\}$ 的子空间, 从而 $Y = \mathbb{R}^n$, 这就蕴涵着 f 在 $\bar{x}$ 处的可微性. 事实上, 任取定 $v \in \partial f(\bar{x})$, 对任意的 $d \in \mathbb{R}^n$, 由定理 2.84(iii) 得 $f'(\bar{x}; d) = \langle v, d\rangle$. 因此 $\partial f(\bar{x})$ 是单点集, 根据定理 3.3 这保证 f 在 $\bar{x}$ 处是可微的.

练习 3.6 利用推论 2.37 并沿用定理 2.40 的证明.

练习 3.7 一个合适的例子由 $\mathbb{R}$ 上的函数 $f(x) := x^4$ 给出.

练习 3.10 任取定序列 $\{x_k\} \subset \operatorname{co}\Omega$, 根据 Carathéodory 定理, 可找到 $\lambda_{k,i} \geqslant 0$ 和 $w_{k,i} \in \Omega, i = 0, \cdots, n$ 满足

$$x_k = \sum_{i=0}^{n} \lambda_{k,i} w_{k,i}, \quad \sum_{i=0}^{n} \lambda_{k,i} = 1, \quad \forall\, k \in \mathbb{N}.$$

于是对 $i = 0, \cdots, n$, 利用 $\{\lambda_{k,i}\}_k$ 的有界性和 Ω 的紧性来抽取 $\{x_k\}$ 的子序列收敛于 $\operatorname{co}\Omega$ 的某一元素.

练习 3.16 根据推论 2.18 有 $N(\bar{x}; \Omega_1 \cap \Omega_2) = N(\bar{x}; \Omega_1) + N(\bar{x}; \Omega_2)$. 利用定理 3.16 得

$$\begin{aligned} T(\bar{x}; \Omega_1 \cap \Omega_2) &= [N(\bar{x}; \Omega_1 \cap \Omega_2)]^\circ \\ &= [N(\bar{x}; \Omega_1) + N(\bar{x}; \Omega_2)]^\circ = [N(\bar{x}; \Omega_1)]^\circ \cap [N(\bar{x}; \Omega_2)]^\circ \\ &= T(\bar{x}; \Omega_1) \cap T(\bar{x}; \Omega_2). \end{aligned}$$

练习 3.18 考虑集合 $\Omega := \{(x, y) \in \mathbb{R}^2 \mid y \geqslant x^2\}$.

练习 3.19 任取定 $a, b \in \mathbb{R}^n$ 且 $a \neq b$. 由定理 3.20 知存在 $c \in (a, b)$ 满足

$$f(b) - f(a) = \langle v, b - a\rangle, \quad \text{对某} \ \ v \in \partial f(c),$$

则 $|f(b) - f(a)| \leqslant \|v\| \cdot \|b - a\| \leqslant \ell\|b - a\|$.

练习 3.21 (i) $F_\infty = \{0\} \times \mathbb{R}_+$. (ii)$F_\infty = F$. (iii) $F_\infty = \mathbb{R}_+ \times \mathbb{R}_+$. (iv)$F_\infty = \{(x, y) \mid y \geqslant |x|\}$.

练习 3.23 (i) 利用定义. (ii) 一个可行的条件是 $A_\infty \cap (-B_\infty) = \{0\}$.

练习 3.24 假设 $\mathcal{T}_\Omega^F(x) = 0$, 则存在序列 $t_k \to 0, t_k \geqslant 0$, 且对任意的 $k \in \mathbb{N}, (x + t_k F) \cap \Omega \neq \varnothing$. 选取 $f_k \in F$ 和 $w_k \in \Omega$ 满足 $x + t_k f_k = w_k$. 由于 F 是有界的, 序列 $\{w_k\}$ 收敛于 x, 从而 $x \in \Omega$, 由于 Ω 是闭的. 反向的蕴涵关系是平凡的.

练习 3.25 沿用定理 3.37 和定理 3.38 的证明.

第 4 章练习

练习 4.2 根据命题 4.4 集合 $\mathcal{U}_\lambda = \mathcal{L}_\lambda^c$ 是闭的.

练习 4.3 考虑 l.s.c. 函数 $f(x) := 0, x \neq 0$ 且 $f(0) := -1$, 则乘积 $f \cdot f$ 不是 l.s.c. 的.

练习 4.7 解方程 $\nabla g(x) = 0$, 其中 $\nabla g(x) = \sum_{i=1}^{m} 2(x - a_i)$.

练习 4.8 设 $\Omega := \{x \in \mathbb{R}^n \mid Ax = b\}$, 则对 $x \in \Omega, N(x;\Omega) = \{A^{\mathrm{T}}\lambda \mid \lambda \in \mathbb{R}^P\}$. 由推论 4.15 知 $x \in \Omega$ 是该问题的最优解当且仅当存在 $\lambda \in \mathbb{R}^P$ 使得

$$-2x = A^{\mathrm{T}}\lambda, \quad Ax = b.$$

解这个方程组得 $x = A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}b$.

练习 4.11 Matlab 函数 $\operatorname{sign} x$ 给出函数 $f(x) := \|x\|_1 (\forall x \in \mathbb{R})$ 的一个次梯度. 考虑集合 $\Omega := \{x \in \mathbb{R}^n \mid Ax = b\}$, 则有 $\Pi(x;\Omega) = x + A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}(b - Ax)$, 这可通过在 $u \in \Omega$ 上最小化二次函数 $\|x - u\|^2$ 的过程中导出, 参见练习 4.8.

练习 4.12 如果对某 $k_0 \in \mathbb{N}, 0 \in \partial f(x_{k_0})$, 则 $\alpha_{k_0} = 0$, 因此 $x_k = x_{k_0} \in S$ 且对任意的 $k \geqslant k_0, f(x_k) = \overline{V}$. 于是, 不失一般性, 假设 $0 \notin \partial f(x_k), k \in \mathbb{N}$. 根据命题 4.23 的证明得

$$0 \leqslant \|x_{k+1} - \bar{x}\|^2 \leqslant \|x_1 - \bar{x}\|^2 - 2\sum_{i=1}^{k} \alpha_i [f(x_i) - \overline{V}] + \sum_{i=1}^{k} \alpha_i^2 \|v_i\|^2$$

对任意的 $\bar{x} \in S$ 成立. 在上式中 α_k 用其表达式替换, 得

$$\begin{aligned} 0 \leqslant \|x_{k+1} - \bar{x}\|^2 &\leqslant \|x_1 - \bar{x}\|^2 - 2\sum_{i=1}^{k} \frac{[f(x_i) - \overline{V}]^2}{\|v_i\|^2} + \sum_{i=1}^{k} \frac{[f(x_i) - \overline{V}]^2}{\|v_i\|^2} \\ &= \|x_1 - \bar{x}\|^2 - \sum_{i=1}^{k} \frac{[f(x_i) - \overline{V}]^2}{\|v_i\|^2}, \end{aligned}$$

且

$$\sum_{i=1}^{k} \frac{[f(x_i) - \overline{V}]^2}{\|v_i\|^2} \leqslant \|x_1 - \bar{x}\|^2.$$

由于对任意的 $i, \|v_i\| \leqslant \ell$, 所以

$$\sum_{i=1}^{k} \frac{[f(x_i) - \overline{V}]^2}{\ell^2} \leqslant \|x_1 - \bar{x}\|^2,$$

这蕴涵着

$$\sum_{i=1}^{k} [f(x_i) - \overline{V}]^2 \leqslant \ell^2 \|x_1 - \bar{x}\|^2, \tag{4.66}$$

因此级数 $\sum_{k=1}^{\infty}[f(x_k)-\overline{V}]^2$ 是收敛的. 因此 $f(x_k)\to\overline{V}(k\to\infty)$, 这就证明了 (i) 成立. 根据式 (4.66) 和事实 $V_k\leqslant f(x_j), j\in\{1,\cdots,k\}$ 得

$$k(V_k-\overline{V})^2\leqslant\ell^2\|x_1-\bar{x}\|^2.$$

于是 $0\leqslant V_k-\overline{V}\leqslant\dfrac{\ell\|x_1-\bar{x}\|}{\sqrt{k}}$. 由于 $\bar{x}$ 是在 S 中任意选取的, 因此得 (ii) 成立.

练习 4.17 定义 Matlab 函数

$$f(x):=\frac{\displaystyle\sum_{i=1}^{m}\frac{a_i}{\|x-a_i\|}}{\displaystyle\sum_{i=1}^{m}\frac{1}{\|x-a_i\|}},\quad x\notin\{a_1,\cdots,a_m\}$$

和 $f(x):=x, x\in\{a_1,\cdots,a_m\}$. 利用 FOR 循环 (其中需要加一个循环停止准则) 来找到 $x_{k+1}:=f(x_k)$, 这里 x_0 是初始点.

练习 4.18 利用次微分 Fermat 法则. 分别考虑 m 是奇数和偶数的情形. 在第一种情形下问题有唯一最优解. 在第二种情形下问题有无穷多个解.

练习 4.22 对区间 $\Omega:=[a,b]\subset\mathbb{R}$, 由距离函数的次微分公式得

$$\partial d(\bar{x};\Omega)=\begin{cases}\{0\}, & \bar{x}\in(a,b),\\ [0,1], & \bar{x}=b,\\ \{1\}, & \bar{x}>b,\\ [-1,0], & \bar{x}=a,\\ \{-1\}, & \bar{x}<a.\end{cases}$$

练习 4.23 对 $\Omega:=\mathbb{B}(c;r)\subset\mathbb{R}^n$, 函数 $d(x;\Omega)$ 在 $\bar{x}$ 处的一个次梯度为

$$v(\bar{x}):=\begin{cases}0, & \bar{x}\in\Omega,\\ \dfrac{\bar{x}-c}{\|\bar{x}-c\|}, & \bar{x}\notin\Omega,\end{cases}$$

则函数 f 的一个次梯度可根据推论 2.46 的次微分和法则找到. 注意到在这种情形下规范条件 (2.29) 自动成立.

练习 4.27 利用练习 4.23 提示中的次梯度公式以及命题 2.54. 注意到容易找到一个元素 $i\in I(\bar{x})$, 则函数 $f_i(x):=d(x;\Omega_i)$ 在 $\bar{x}$ 处的任何次梯度都属于 $\partial f(\bar{x})$.

参考文献

[1] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011. DOI: 10.1007/978-1-4419-9467-7.

[2] D. P. Bertsekas, A. Nedić and A. E. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, 2003.

[3] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. DOI: 10.1017/CBO9780511804441. xi

[4] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, Springer, 2000. DOI: 10.1007/978-1-4757-9859-3. xi

[5] J. M. Borwein and Q. J. Zhu, *Techniques of Variational Analysis*, Springer, 2005.

[6] Z. Drezner and H. W. Hamacher (Eds.), *Facility Location: Applications and Theory*, Springer, 2004. 154

[7] G. Colombo and P. R. Wolenski, "Variational analysis for a class of minimal time functions in Hilbert spaces," *J. Convex Anal.*, vol. 11, pp. 335-361, 2004.

[8] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Vols. I and II, Springer, 1993. DOI: 10.1007/978-3-662-06409-2. xi

[9] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, Springer, 2001. DOI: 10.1007/978-3-642-56468-0. xi, 185

[10] H. W. Kuhn, "A note on Fermat-Torricelli problem," *Math. Program.*, vol. 54, pp. 98.107, 1973. DOI: 10.1007/BF01584648. 151

[11] Y. S. Kupitz, H. Martini and M. Spirova, "The Fermat-Torricelli problem, Part I: A discrete gradient-method approach," *Optim. Theory Appl.*, vol. 158, pp. 305.327, 2013. DOI: 10.1007/s10957-013-0266-z. 154

[12] S. R. Lay, *Convex Sets and Their Applications*, Wiley & Sons, 1982.

[13] G. G. Magaril-Il'yaev and V. M. Tikhomirov, *Convex Analysis: Theory and Applications*, American Mathematical Society, 2003. 196 BIBLIOGRAPHY

[14] B. S. Mordukhovich, *Variational Analayis and Generalized Differentiation*, Vols. I and II, Springer, 2006. DOI: 10.1007/3-540-31246-3. 40, 41, 49

[15] B. S. Mordukhovich and N. M. Nam, "Applications of variational analysis to a generalized Fermat-Torricelli problem," *Optim. Theory Appl.*, vol. 148, pp. 431.454, 2011. DOI: 10.1007/s10957-010-9761-7. 154

[16] B. S. Mordukhovich, N. M. Nam and J. Salinas, "Solving a generalized Heron problem by means of convex analysis," *Amer. Math. Monthly*, vol. 119, pp. 87.99, 2012. DOI: 10.4169/amer.math.monthly.119.02.087. 154

[17] B. S. Mordukhovich, N. M. Nam and J. Salinas, "Applications of variational analysis to a generalized Heron problem," *Applic. Anal.*, vol. 91, pp. 1915.1942, 2012. DOI: 10.1080/00036811.2011.604849. 154

[18] B. S. Mordukhovich, N. M. Nam and C. Villalobos, "The smallest enclosing ball problem and the smallest intersecting ball problem: Existence and uniqueness of solutions," *Optim. Letters*, vol. 154, pp. 768-791, 2012. DOI: 10.1007/s11590-012-0483-7. 168

[19] N. M. Nam and N. Hoang, "A generalized Sylvester problem and a generalized Fermat-Torricelli problem," *J. Convex Anal.*, vol. 20, pp. 669-687, 2013. 168

[20] N. M. Nam, N. Hoang and N. T. An, "Constructions of solutions to generalized Sylvester and Fermat-Torricelli problems for Euclidean balls," *Optim. Theory Appl.*, in press. DOI: 10.1007/s10957-013-0366-9. 168

[21] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, 2004. DOI: 10.1007/978-1-4419-8853-9. xi

[22] Y. Nesterov, "Smooth minimization of nonsmooth functions," *Math Program.*, vol. 103, pp. 127.152, 2005. DOI: 10.1007/s10107-004-0552-5. 191

[23] J. F. Nash, *Noncooperative Games*, Doctoral thesis, Princeton University, 1950. 122

[24] J. F. Nash, "Equilibrium points in N-person games," *Proc. Nat. Acad. Sci.* vol. 36, pp. 48-49, 1950. DOI: 10.1073/pnas.36.1.48. 122

[25] B. T. Polyak, *An Introduction to Optimization*, Optimization Solftware, 1987.

[26] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970. xi, 54

[27] R. T. Rockafellar and R. J-B. Wets, *Variational Analysis*, Springer, 1998. DOI: 10.1007/978- 3-642-02431-3. 49 BIBLIOGRAPHY 197

[28] R. T. Rockafellar, "Applications of convex variational analysis to Nash equilibrium," *Proceedings of 7th International Conference on Nonlinear Analysis and Convex Analysis* (Busan,Korea), pp. 173-183, 2011. 123

[29] A. Ruszczynski, *Nonlinear Optimization*, Princeton University Press, 2006. 154

[30] N. Z. Shor, *Minimization Methods for Nondifferentiable Functions*, Springer, 1985. DOI: 10.1007/978-3-642-82118-9.

[31] J. J. Sylvester, "A question in the geometry of situation," *Quart. J. Math.*, vol. 1, p. 79, 1857. 168

[32] E. Weiszfeld, "Sur le point pour lequel la somme des distances de n points donn é s est minimum," *Tôohoku Mathematics Journal*, vol. 43, pp. 355 - 386, 1937. 150, 151

[33] E. Zeidler, *Applied Functional Analysis. Applications to Mathematical Physics*, Springer, 1997. 123

索　引

(O-6126.31)

www.sciencep.com



科学数理分社
电话：(010)64011058
E-mail：lixin_kx@mail.sciencep.com

销售分类建议：高等数学

定 价：78.00 元